Predicting the Effectiveness of a Mindfulness Virtual Community Intervention for University Students: Machine Learning Model (e50982)
Christo El Morr, Farideh Tavangar, Farah Ahmad, Paul Ritvo, MVC Team ................................................................. 5

Improvements in Neoplasm Classification in the International Classification of Diseases, Eleventh Revision: Systematic Comparative Study With the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (e52296)
Yicong Xu, Jingya Zhou, Hongxia Li, Dong Cai, Huanbing Zhu, Shengdong Pan ................................................................. 77

Evaluation of an mHealth App on Self-Management of Osteoporosis: Prospective Survey Study (e53995)
Magnus Bendtsen, Bodil Schönwandt, Mette Rubæk, Mette Hitz .................................................................................. 109

Intramural Health Care Through Video Consultations and the Need for Referrals and Hospital Admissions: Retrospective Quantitative Subanalysis of an Evaluation Study (e44906)
Katharina Schmalstieg-Bahr, Miriam Colombo, Roland Koch, Joachim Szecsenyi, Friedrich Völker, Eva Blozik, Martin Scherer .................................................... 120

Evolution of Digital Health and Exploration of Patented Technologies (2017-2021): Bibliometric Analysis (e48259)
Wenjun Gu, Jinhua Wang, Yunqi Zhang, Shaolin Liang, Zisheng Ai, Jiyu Li ........................................................................ 130

Behavioral Insights from Vaccine Adoption in Nigeria: Cross-Sectional Survey Findings (e47817)
Sohail Agha, Ifeanyi Nsofor, Drew Bernard, Sarah Francis, Nandan Rao ................................................................. 149

Influence of Environmental Factors and Genome Diversity on Cumulative COVID-19 Cases in the Highland Region of China: Comparative Correlational Study (e43585)
Zhuoga Deji, Yuantao Tong, Honglian Huang, Zeyu Zhang, Meng Fang, M Crabbe, Xiaoyan Zhang, Ying Wang ................................................................. 161

The Association Between Depressive Symptoms and the Weekly Duration of Physical Activity Subset by Intensity and Domain: Population-Based, Cross-Sectional Analysis of the National Health and Nutrition Examination Survey From 2007 to 2018 (e48396)
The Effect of Body Temperature Changes on the Course of Treatment in Patients With Pneumonia and Sepsis: Results of an Observational Study (e52590)
Domen Guzelj, Anže Grubelnik, Nina Greif, Petra Povalej Bržan, Jure Fluher, Žiga Kalamar, Andrej Markota... 189

Evaluation of the Accuracy, Credibility, and Readability of Statin-Related Websites: Cross-Sectional Study (e42849)
Eunice Ling, Domenico de Pieri, Evenne Loh, Karen Scott, Stephen Li, Heather Medbury... 218

Online Visibility and Scientific Relevance of Strabismus Research: Bibliometric Analysis (e50698)
Aleksander Stupnicki, Basil Suresh, Saurabh Jain... 230

Census-Dependent Mortality of Ventilated Patients With COVID-19 in Israel: Noninterventional Observational Cohort Study (e41749)
Joseph Mendlovic, Francis Mimouni... 235

Chinese Health Insurance in the Digital Era: Bibliometric Study (e52020)
Zhiyuan Hu, Xiaopeng Qin, Kaiyan Chen, Yu-Ni Huang, Richard Wang, Tao-Hsin Tung, Yen-Ching Chuang, Bing-Long Wang... 245

Biochemical Changes in Adult Male Gamers During Prolonged Gaming: Pilot Study (e46570)
Kasper Krarup, Johannes Riis, Morten Mørk, Hien Nguyen, Inge Søkilde Pedersen, Søren Risom Kristensen, Aase Handberg, Henrik Krarup... 258

Effectiveness of a Smartphone App to Promote Physical Activity Among Persons With Type 2 Diabetes: Randomized Controlled Trial (e53054)
Stephanie Bonn, Madeleine Hummel, Giulia Peveri, Helén Eke, Christina Alexandrou, Rino Bellocco, Marie Löf, Ylva Trolle Lagerros... 269

A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence–Based Models in Health Care Education and Practice: Development Study Involving a Literature Review (e54704)
Malik Sallam, Muna Barakat, Mohammed Sallam... 283

Research Trends and Evolution in Radiogenomics (2005-2023): Bibliometric Analysis (e51347)
Meng Wang, Yun Peng, Ya Wang, Dehong Luo... 304

Cardiovascular Comorbidities in COVID-19: Comprehensive Analysis of Key Topics (e55699)
Rene Markovi, Luka Ternar, Tim Trstenjak, Marko Marhl, Vladimir Grubelnik... 317

Factors Associated With Worsened Mental Health of Health Care Workers in Canada During the COVID-19 Pandemic: Cross-Sectional Survey Study (e50064)
Ali Abdul-Hussein, Zahid Butt, Stanko Dimitrov, Brian Cozzarin... 332

A Web-Based Tool to Assess Social Inclusion and Support Care Planning in Mental Health Supported Accommodation: Development and Preliminary Test Study (e45987)
Sharon Eager, Helen Killaspy, Joanna C, Gillian Mezey, Peter McPherson, Megan Downey, Georgina Thompson, Brynmor Lloyd-Evans... 361

Patterns of Skin Picking in Skin Picking Disorder: Ecological Momentary Assessment Study (e53831)
Christina Gallinat, Markus Moessner, Maximilian Wilhelm, Nancy Keuthen, Stephanie Bauer... 361

Effects of Background Music on Attentional Networks of Children With and Without Attention Deficit/Hyperactivity Disorder: Case Control Experimental Study (e53869)
Camila Mendes, Jonas de Paula, Débora Miranda... 373
Viewpoints

Designing mHealth Apps to Incorporate Evidence-Based Techniques for Prolonging User Engagement (e51974)
Rebecca Monachelli, Sharon Davis, Allison Barnard, Michelle Longmire, John Docherty, Ingrid Oakley-Girvan. .......................................................... 19

Using Routine Data to Improve Lesbian, Gay, Bisexual, and Transgender Health (e53311)
Catherine Saunders. .................................................................................................................. 27

Ad Hoc Modifications to a High Dependency Psychiatric Unit for People With Dementia During the COVID-19 Period (e49618)
Thanita Pilunthanakul, Giles Tan. .......................................................................................... 36

Reviews

Strategies to Alleviate the Burden Experienced by Informal Caregivers of Persons With Severe Mental Disorders in Low- and Middle-Income Countries: Scoping Review (e48587)
Olindah Silaule, Daleen Casteleijn, Fasloen Adams, Nokuthula Nkosi. ........................................ 40

Digital Methods for the Spiritual and Mental Health of Generation Z: Scoping Review (e48929)
Susanna Park, Bridgette Do, Jaconlyn Yourell, Janice Hermer, Jennifer Huberty. .......................... 65

Theoretical Perspectives Underpinning Research on the Physician-Patient Relationship in a Digital Health Practice: Scoping Review (e47280)
Damoun Nassehi, Birgitta Gripsrud, Ellen Ramvi. ......................................................................... 98

Gamification and Oral Health in Children and Adolescents: Scoping Review (e35132)
Rui Moreira, Augusta Silveira, Teresa Sequeira, Nuno Durão, Jessica Lourenço, Inês Cascais, Rita Cabral, Tiago Taveira Gomes. .................................. 205

Original

Application of AI in Sepsis: Citation Network Analysis and Evidence Synthesis (e54490)
MeiJung Wu, Md Islam, Tahmina Poly, Ming-Chin Lin. .................................................................. 87

Short Paper

Validation and Refinement of the Sense of Coherence Scale for a French Population: Observational Study (e50284)
Paul Sebo, Benoit Tudrej, Augustin Bernard, Bruno Delaunay, Alexandra Dupuy, Claire Malavergne, Hubert Maisonneuve. ........................................ 199

Corrigenda and Addenda

Correction: Influence of Environmental Factors and Genome Diversity on Cumulative COVID-19 Cases in the Highland Region of China: Comparative Correlational Study (e59070)
Zhuoga Deji, Yuantao Tong, Honglian Huang, Zeyu Zhang, Meng Fang, M Crabbe, Xiaoyan Zhang, Ying Wang. .................................................. 315
Case Reports

Treating Spontaneous Pneumothorax Using an Innovative Surgical Technique Called Capnodissection Pleurectomy: Case Report (e54497)
Ghaith Osous, Prashanth Ramaraj, Sanjeet Avtaar Singh, Philip Herd, Nayandra Sooraj, Malcolm Will. ................................. 386

Radiological Progression of Degenerative Cervical Myelopathy in a Clinically Stable Patient: Case Report (e48212)
Rishi Umeria, Oliver Mowforth, Munashe Veremu, Benjamin Davies, Mark Kotter. ............................................................. 390

Research Letter

Visual “Scrollytelling”: Mapping Aquatic Selfie-Related Incidents in Australia (e53067)
Samuel Cornell, Amy Peden. .................................................................................................................. 397
Predicting the Effectiveness of a Mindfulness Virtual Community Intervention for University Students: Machine Learning Model

Christo El Morr¹, PhD; Farideh Tavangar², MSc; Farah Ahmad¹, PhD; Paul Ritvo³, PhD; MVC Team⁴

¹School of Health Policy and Management, York University, Toronto, ON, Canada
²Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada
³Kinesiology & Health Science, York University, Toronto, ON, Canada
⁴See Acknowledgments

Abstract

Background: Students’ mental health crisis was recognized before the COVID-19 pandemic. Mindfulness virtual community (MVC), an 8-week web-based mindfulness and cognitive behavioral therapy program, has proven to be an effective web-based program to reduce symptoms of depression, anxiety, and stress. Predicting the success of MVC before a student enrolls in the program is essential to advise students accordingly.

Objective: The objectives of this study were to investigate (1) whether we can predict MVC’s effectiveness using sociodemographic and self-reported features and (2) whether exposure to mindfulness videos is highly predictive of the intervention’s success.

Methods: Machine learning models were developed to predict MVC’s effectiveness, defined as success in reducing symptoms of depression, anxiety, and stress as measured using the Patient Health Questionnaire-9 (PHQ-9), the Beck Anxiety Inventory (BAI), and the Perceived Stress Scale (PSS), to at least the minimal clinically important difference. A data set representing a sample of undergraduate students (N=209) who took the MVC intervention between fall 2017 and fall 2018 was used for this secondary analysis. Random forest was used to measure the features’ importance.

Results: Gradient boosting achieved the best performance both in terms of area under the curve (AUC) and accuracy for predicting PHQ-9 (AUC=0.85 and accuracy=0.83) and PSS (AUC=1 and accuracy=1), and random forest had the best performance for predicting BAI (AUC=0.93 and accuracy=0.93). Exposure to online mindfulness videos was the most important predictor for the intervention’s effectiveness for PHQ-9, BAI, and PSS, followed by the number of working hours per week.

Conclusions: The performance of the models to predict MVC intervention effectiveness for depression, anxiety, and stress is high. These models might be helpful for professionals to advise students early enough on taking the intervention or choosing other alternatives. The students’ exposure to online mindfulness videos is the most important predictor for the effectiveness of the MVC intervention.

Trial Registration: ISRCTN Registry ISRCTN12249616; https://www.isrctn.com/ISRCTN12249616

doi:10.2196/50982

KEYWORDS

machine learning; virtual community; virtual care; mindfulness; depression; anxiety; stress; students; online; randomized controlled trial; Canada; virtual; artificial intelligence; symptoms; behavioral therapy; sociodemographic; mindfulness video; online video
Students’ mental health crises were recognized before the COVID-19 pandemic and deepened during the pandemic. University students are experiencing an increase in psychological distress on North American campuses. A student survey of 32 Canadian postsecondary institutions reported high anxiety (56.5%), hopelessness (54%), seriously depressed mood (37.5%), and overwhelming anger (42%) [1]. A similar survey in 2016 revealed higher distress levels [2]. In 2013, a study of 997 students at York University (site of this study) indicated that 57% reported depression scores sufficient for diagnosable clinical depression, while 33% reported anxiety scores in ranges typically indicative of panic disorder and generalized anxiety disorder [3]. The situation appears similar at universities in the United States [4,5] and worldwide; in 2018, the World Health Organization reported increasing mental disorders in college and university students worldwide [6]. Mental health challenges among university students demand attention. This is a vulnerable period, as 70% of mental health problems emerge before the age of 25 years. Without intervention, these problems can worsen and hinder students’ personal and academic success [7]. COVID-19 has negatively impacted university students’ mental health [8-10].

University student distress is both an individual and societal challenge. Losses in productivity during the study and at work due to distress and mental disorders are associated with indirect but significant economic burdens [11]. Canadian estimates show that mental disorders cost nearly US $37 billion yearly, with 9.8% due to direct medical costs, 16.6% and 18.2% due to long-term loss and short-term work loss, respectively, and 55.4% due to the loss of healthy function (ie, loss of the utilities of vision, hearing, speech, mobility, dexterity, emotion, cognition, and pain as assessed in the Health Utilities Index Mark 3 system) [12].

While mental distress and disorder are becoming more prevalent in students, the counseling offered in colleges and universities needs to catch up with demand. For example, from 2007 to 2012, full-time enrollment in the Ontario college system increased from 167,000 to 210,600 (a 26% increase), while the number of counselors employed in the college system increased from 146 to 152.7 (a 4.6% increase) [13]. This discrepancy leaves students underserved and counselors overwhelmed amid the increasing distress [14].

Mindfulness-based interventions have been demonstrated to positively impact psychological and physical health [15-17], with multiple meta-analyses demonstrating positive impacts in clinical and nonclinical populations [18-22]. However, with large numbers of students (50,000 to 60,000 on some campuses), there may not be enough trained personnel to convey helpful mindfulness-based practices directly. Instead, in the eHealth domain, virtual communities (VCs) [23], that is, online mindfulness-based practices, have been used in health care to provide e-education tools and online support to empower active participants in health enhancement [24-26]. VCs can scale up mindfulness interventions at lower costs to a broader range of students, especially those restricted from attending clinics due to time-place discontinuities. VCs preserve anonymity (with reduced stigmatization) while promoting voluntary, supportive, interpersonal connections.

We developed a web-delivered mindfulness program (mindfulness virtual community [MVC]) to reduce symptoms of depression, anxiety, and stress in university students and conducted a randomized controlled trial (RCT) targeting university students at a Canadian university to examine its effectiveness. Following a successful RCT [26-29], we wanted in this secondary analysis (1) to develop a machine learning (ML) model to predict the effectiveness of the online mindfulness intervention on mental health outcomes using sociodemographic and self-reported features and (2) to investigate if exposure to mindfulness videos was highly predictive of the intervention’s success.

**Methods**

**Prediction Problem**

This study aims to predict the effectiveness (ie, success vs nonsuccess) of the online mindfulness intervention on mental health outcomes; as such, this is a retrospective prognostic analysis of a classification problem per individual (ie, participants in the MVC mindfulness intervention).

**Data Set Source**

This is a retrospective analysis, where we analyzed an anonymized data set. The data were deidentified, and consent was obtained during the RCT; no further consent was sought for this secondary data analysis since nonidentifiable data were used. The data set was collected via an RCT described in detail elsewhere [28]. The parent study design consisted of a 2-arm parallel-design RCT, comparing a group assigned to the web-based MVC program to a waitlist control group. Participants in the study were students who were at least 18 years of age, reported English language fluency, self-reported high confidence in completing the study, and actively enrolled in an undergraduate program. This paper is based on the MVC intervention sample recruited in fall 2017, winter 2017, and fall 2018. The MVC intervention was an 8-week program and was comprised of three components: (1) 12 online videos for mental health education; (2) 3 anonymous discussion boards on depression, anxiety, and stress; and (3) anonymous, 20-minute group-based live videoconferences led by a mental health professional with training in mindfulness during which students could raise questions related to mindfulness (Figure 1).

Each of the 12 mental health modules consisted of 1 educational content video and 1 mindfulness practice video recorded in both male and female voices and offered in both high and low resolution (a total of 8 videos per module); participants could choose the type of video they wanted to watch for each module. The videos were available for participants 24 hours a day to watch or listen to on computers, phones, or tablets at their convenience. The module scripts and audio recordings were created by one of the investigators with extensive experience as a clinical psychologist and researcher in mindfulness. They were based on mindfulness and cognitive behavioral therapy principles and informed by the prior student-based focus group...
study [30,31]—the choice of moving and still images used in creating the videos involved collaborative work. The topics of the 12 modules included the following: overcoming stress, anxiety, and depression; mindfulness and being a student; mindfulness for better sleep; thriving in a fast-changing world; healthy intimacy; destigmatization; no more procrastination; pain reduction and mindfulness; healthy body image; healthier eating; overcoming trauma; and relationships with family and friends.

The primary RCT outcomes were depression, anxiety, and perceived stress, following hypotheses that symptom scores for depression, anxiety, and stress at T2 (after 8 weeks) would be significantly better in the MVC group when compared with the waitlist control group. The outcomes were measured with the following validated scales: Patient Health Questionnaire-9 (PHQ-9) [32], Beck Anxiety Inventory (BAI) [33], and Perceived Stress Scale (PSS) [34]. The secondary aim was to assess the impact of 3 elements of the MVC intervention on the outcomes. Participants also completed a sociodemographic questionnaire section at the T1 (baseline) survey.

**Ethical Considerations**

The previous study received ethics approval from the Human Participant Research Committee (certificate e2016-345) of the York University. This ML study received ethics approval from the same committee (certificate e2023-012); the approval covers secondary analysis without additional consent. Participants in the original study had the option to receive an honorarium of CAD $50 (US $37.5) or 2% in course grade (for professors who gave this permission) or 3 credits (equivalent to 2% course grade) in the Undergraduate Research Participation Pool of the Department of Psychology. The participants’ data were anonymized.

**Participants**

We aimed to build a model to predict who will likely benefit from the intervention, unlike the RCT study, where overall intervention effectiveness was determined (and supported by analysis) by comparing intervention and control groups. That is why we have analyzed intervention group data only to understand individual differences in response to the intervention.

**Data Preparation**

The data set consisted of 209 students who took the MVC intervention during fall 2017, winter 2018, and fall 2018. The effectiveness of the intervention was determined using the minimal clinically important difference (MCID), that is, the level of reduction in symptoms that psychologists consider clinically meaningful, for each of the mental health outcomes. We adopted evidence from psychology that determines the MCID to be a 5-point reduction in PHQ-9 for depression [35,36], an 8.8-point reduction in BAI for anxiety [30,31], and an 11-point reduction in PSS for stress [37,38]. Any reduction equal to or above the MCID was labeled an effective intervention (label=1); otherwise, it was deemed ineffective (label=0).

To build a good prediction model from the training set, the data must be balanced. The class labels of the target variables, PHQ-9, BAI, and PSS, used in this study were not balanced. In our case, the percentage of instances with label=1 was extremely low: 50 (23%) for PHQ-9, 48 (24%) for BAI, and 8 (3.8%) for PSS, leading to a substantial imbalance. To alleviate the imbalanced data, we applied an oversampling method using the sklearn.resample function available in Python (version 3; Python Software Foundation).

**Missing Data**

Missing data in the outcomes were 12 (5.7%) for BAI and PHQ-9 and 13 (6.2%) for PSS of the 209 records. Missing data for the outcomes were dropped from the data set. There were no missing values in the predictors.

**Labels and Features**

The outcome variables were the 3 MCIDs associated with PHQ-9, BAI, and PSS being met or not for each instance. To investigate whether we can predict MVC’s effectiveness using sociodemographic and self-reported features, the following features were used: sex (male and female), country of birth...
(Canada and other), first language (English and other), education level (bachelor degree and other), ethnicity (White and non-White), marital status (married and other), age, number of weekly working hours, and self-rated health (poor, fair, good, very good, and excellent). To investigate the importance of exposure to mindfulness videos, in comparison with these features, in the prediction of intervention success, we added the total number of mindfulness videos watched to the previous data set.

**Algorithms**

Seven different classification algorithms, representing different learning paradigms, were used in this study: logistic regression (LR), support vector machine (SVM), random forest (RF), decision tree (DT), k-nearest neighbor (KNN), adaptive boosting (AdaBoost), and gradient boosting that showed good performance in previous studies that targeted depression, anxiety, and stress [35,39,40]. The implementations of the classification algorithms provided in the scikit-learn ML library [41] were used. The data set was split into 80% for training and 20% for testing. Hyperparameter tuning for each algorithm was performed using a grid search over a 10-fold cross-validation on the training data set. The optimal hyperparameters for the classification algorithms and their values for the data set without exposure to videos and the data set with exposure to videos are presented in Tables 1 and 2, respectively.

Each classifier’s performance was compared with the best overall performance, leading to the selection of the best prediction model for the psychological outcomes. The classifiers’ performances were assessed based on several evaluation metrics, including the percentage of correctly classified instances or the accuracy, sensitivity, specificity, and area under the curve (AUC) of the receiver operating characteristic curve. The best performance, as measured by the AUC score, was chosen for each algorithm.

To evaluate the features’ importance in predicting intervention success, the data set with the total exposure to mindfulness videos was used to build predictive models. The RF algorithm was used to measure the features’ importance. The hyperparameters used for the classification algorithms and their values that provided the optimal model are presented in Table 2.
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>C=1, penalty=l1, solver=liblinear</td>
</tr>
<tr>
<td>BAI</td>
<td>C=1, penalty=l1, solver=liblinear</td>
</tr>
<tr>
<td>PSS</td>
<td>C=1, penalty=l1, solver=liblinear</td>
</tr>
<tr>
<td>Support vector machine</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>C=10, γ=0.1, kernel=rbf</td>
</tr>
<tr>
<td>BAI</td>
<td>C=10, γ=0.1, kernel=rbf</td>
</tr>
<tr>
<td>PSS</td>
<td>C=10, γ=0.1, kernel=rbf</td>
</tr>
<tr>
<td>Random forest</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>Max_features=auto, n_estimators=500, max_depth=8, criterion=entropy</td>
</tr>
<tr>
<td>BAI</td>
<td>Max_features=auto, n_estimators=500, max_depth=8, criterion=gini</td>
</tr>
<tr>
<td>PSS</td>
<td>Max_features=auto, n_estimators=500, max_depth=8, criterion=gini</td>
</tr>
<tr>
<td>Decision tree</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>Max_leaf_nodes=59, random_state=42, min_samples_split=2, criterion=entropy</td>
</tr>
<tr>
<td>BAI</td>
<td>Max_leaf_nodes=56, random_state=42, min_samples_split=2, criterion=entropy</td>
</tr>
<tr>
<td>PSS</td>
<td>Max_leaf_nodes=16, random_state=42, min_samples_split=2, criterion=entropy</td>
</tr>
<tr>
<td>K-nearest neighbor</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>N_neighbors=2, weight=distance, leaf size=27, P=1</td>
</tr>
<tr>
<td>BAI</td>
<td>N_neighbors=2, weight=distance, leaf size=1, P=1</td>
</tr>
<tr>
<td>PSS</td>
<td>N_neighbors=1, weight=dniform, leaf size=1, P=1</td>
</tr>
<tr>
<td>Adaptive boosting</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>n_estimators=5000, max_depth=3, learning rate=0.5</td>
</tr>
<tr>
<td>BAI</td>
<td>n_estimators=5000, max_depth=3, learning rate=0.9</td>
</tr>
<tr>
<td>PSS</td>
<td>n_estimators=500, max_depth=3, learning rate=0.9</td>
</tr>
<tr>
<td>Gradient boosting</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>Learning rate=0.05, max depth=6, n_estimators=100, subsample=0.9, max_features=none, min_samples_split=2</td>
</tr>
<tr>
<td>BAI</td>
<td>Learning rate=0.02, max depth=10, n_estimators=1000, subsample=1.0, max_features=none, min_samples_split=2</td>
</tr>
<tr>
<td>PSS</td>
<td>Learning rate=0.01, max depth=6, n_estimators=1000, subsample=0.9, max_features=sqrt, min_samples_split=2</td>
</tr>
</tbody>
</table>

aPHQ-9: Patient Health Questionnaire-9.  
bBAI: Beck Anxiety Inventory.  
cPSS: Perceived Stress Scale.
Table 2. Algorithms and their corresponding optimal hyperparameters found by grid search (data set with exposure to videos).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>C=1, penalty=l1, solver=liblinear</td>
</tr>
<tr>
<td>BAI</td>
<td>C=0.1, penalty=l2, solver=newton-cg</td>
</tr>
<tr>
<td>PSS</td>
<td>C=100, penalty=l2, solver=lbfgs</td>
</tr>
<tr>
<td>Support vector machine</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>C=10, γ=0.01, kernel=rbf</td>
</tr>
<tr>
<td>BAI</td>
<td>C=1, γ=1, kernel=rbf</td>
</tr>
<tr>
<td>PSS</td>
<td>C=1, γ=1, kernel=rbf</td>
</tr>
<tr>
<td>Random forest</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>Max_features=auto, n_estimators=500, max_depth=7, criterion=entropy</td>
</tr>
<tr>
<td>BAI</td>
<td>Max_features=auto, n_estimators=200, max_depth=8, criterion=gini</td>
</tr>
<tr>
<td>PSS</td>
<td>Max_features=auto, n_estimators=500, max_depth=8, criterion=gini</td>
</tr>
<tr>
<td>Decision tree</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>Max_leaf_nodes=39, random_state=42, min_samples_split=2, criterion=entropy</td>
</tr>
<tr>
<td>BAI</td>
<td>Max_leaf_nodes=53, random_state=42, min_samples_split=3, criterion=gini</td>
</tr>
<tr>
<td>PSS</td>
<td>Max_leaf_nodes=16, random_state=42, min_samples_split=2, criterion=gini</td>
</tr>
<tr>
<td>K-nearest neighbor</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>N_neighbors=1, weight=uniform, leaf size=14, P=1</td>
</tr>
<tr>
<td>BAI</td>
<td>N_neighbors=2, weight=distance, leaf size=1, P=1</td>
</tr>
<tr>
<td>PSS</td>
<td>N_neighbors=2, weight=uniform, leaf size=1, P=1</td>
</tr>
<tr>
<td>Adaptive boosting</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>n_estimators=500, max_depth=3, learning rate=0.5</td>
</tr>
<tr>
<td>BAI</td>
<td>n_estimators=500, max_depth=3, learning rate=0.7</td>
</tr>
<tr>
<td>PSS</td>
<td>n_estimators=2000, max_depth=3, learning rate=0.7</td>
</tr>
<tr>
<td>Gradient boosting</td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>Learning rate=0.5, max depth=50, n_estimators=50, subsample=0.9, max_features=sqrt, min_samples_split=2</td>
</tr>
<tr>
<td>BAI</td>
<td>Learning rate=0.04, max depth=10, n_estimators=1000, subsample=0.5, max_features=none, min_samples_split=2</td>
</tr>
<tr>
<td>PSS</td>
<td>Learning rate=0.03, max depth=8, n_estimators=1000, subsample=0.5, max_features=none, min_samples_split=2</td>
</tr>
</tbody>
</table>

*aPHQ-9: Patient Health Questionnaire-9.
*bBAI: Beck Anxiety Inventory.
*cPSS: Perceived Stress Scale.

Results

Demographics

Table 3 presents the demographic characteristics of participants at baseline. Of 209 students, 73.2% (n=153) were female, 8.1% (n=17) were married, and 21.1% (n=44) were White. Most participants were born in Canada, and English was their first language. The median (IQR) of age, work hours per week, and the total number of mindfulness videos watched were 21 (19-23) years, 10 (0-18), and 16 (9-30), respectively.
Objective 1: Predicting MVC’s Effectiveness Using Sociodemographic and Self-Reported Features

Table 4 summarizes the evaluated models’ performances: sensitivity, specificity, accuracy, and AUC, using 10-fold cross-validation.

The results showed that both gradient boosting (AUC=0.85 and accuracy=0.83) and DT (AUC=0.84 and accuracy=0.81) are slightly better compared to AdaBoost and KNN (AUC=0.82 and accuracy=0.80) as well as SVM (AUC=0.81 and accuracy=0.80) and outperformed the remaining classification algorithms for predicting a clinically significant reduction in PHQ-9. The best classifiers for predicting a clinically significant reduction in BAI were RF (AUC=0.93 and accuracy=0.93), followed by gradient boosting (AUC=0.92 and accuracy=0.92) and AdaBoost (AUC=0.91 and accuracy=0.92) and gradient boosting (AUC=0.91 and accuracy=0.92), which outperformed the remaining classifiers. Four classifiers, gradient boosting, AdaBoost, RF, and SVM, gained the perfect AUC and accuracy (AUC=1 and accuracy=1) for predicting a clinically significant reduction in PSS, followed by the near-perfect scores for KNN (AUC=0.99 and accuracy=0.99) and DT (AUC=0.97 and accuracy=0.97). Meanwhile, LR had the lowest performance for PHQ-9, BAI, and PSS in terms of AUC (0.64, 0.75, and 0.73, respectively) and accuracy (0.66, 0.75, and 0.73, respectively).

The results were close to those found in the models built without video exposure. Gradient boosting (AUC=0.89 and accuracy=0.88) was the best predictor for a significant reduction in PHQ-9, followed closely by AdaBoost and DT (AUC=0.84 and accuracy=0.81), which outperformed the remaining classification algorithms. The best classifiers for predicting a clinically significant reduction in BAI were AdaBoost and SVM (AUC=0.93 and accuracy=0.93), followed closely by gradient boosting (AUC=0.93 and accuracy=0.92) and RF (AUC=0.90 and accuracy=0.90), which outperformed the remaining classifiers. Four classifiers, gradient boosting, AdaBoost, RF, and SVM, gained the perfect AUC and accuracy (AUC=1 and accuracy=1) for predicting a clinically significant reduction in PSS, followed by the near-perfect score for KNN (AUC=0.99 and accuracy=0.99) and DT (AUC=0.97 and accuracy=0.97). Meanwhile, LR had the lowest performance for PHQ-9, BAI, and PSS in terms of AUC (0.62, 0.60, and 0.79, respectively) and accuracy (0.63, 0.60, and 0.80, respectively).

Using the second data set (ie, enriched with the exposure to videos), RF was used to detect features’ importance in relation
to the 3 outcomes. The most predictive feature for the PHQ-9, BAI, and PSS was the total exposure to the mindfulness videos, followed by the average number of working hours per week and age for PHQ-9 and BAI. In contrast, age and the average number of working hours per week were the second and third most important predictors for PSS, respectively.

### Table 4. Classification report of the machine learning algorithms for outcomes.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>AUC(^a)</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Logistic regression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9(^b)</td>
<td>0.64</td>
<td>0.66</td>
<td>0.57</td>
<td>0.72</td>
</tr>
<tr>
<td>BAI</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>PSS(^d)</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.74</td>
</tr>
<tr>
<td><strong>Support vector machine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.81</td>
<td>0.80</td>
<td>0.90</td>
<td>0.75</td>
</tr>
<tr>
<td>BAI</td>
<td>0.77</td>
<td>0.77</td>
<td>0.79</td>
<td>0.75</td>
</tr>
<tr>
<td>PSS</td>
<td>0.96</td>
<td>0.96</td>
<td>1.0</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>Random forest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.78</td>
<td>0.76</td>
<td>0.87</td>
<td>0.69</td>
</tr>
<tr>
<td>BAI</td>
<td>0.93</td>
<td>0.93</td>
<td>0.86</td>
<td>1.0</td>
</tr>
<tr>
<td>PSS</td>
<td>0.99</td>
<td>0.99</td>
<td>1.0</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>Decision tree</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.84</td>
<td>0.81</td>
<td>0.96</td>
<td>0.72</td>
</tr>
<tr>
<td>BAI</td>
<td>0.84</td>
<td>0.83</td>
<td>0.93</td>
<td>0.75</td>
</tr>
<tr>
<td>PSS</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>K-nearest neighbor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.82</td>
<td>0.80</td>
<td>0.91</td>
<td>0.72</td>
</tr>
<tr>
<td>BAI</td>
<td>0.78</td>
<td>0.78</td>
<td>0.68</td>
<td>0.88</td>
</tr>
<tr>
<td>PSS</td>
<td>0.96</td>
<td>0.96</td>
<td>1.0</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>Adaptive boosting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.82</td>
<td>0.80</td>
<td>0.91</td>
<td>0.72</td>
</tr>
<tr>
<td>BAI</td>
<td>0.92</td>
<td>0.92</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>PSS</td>
<td>0.99</td>
<td>0.99</td>
<td>1.0</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>Gradient boosting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.85</td>
<td>0.83</td>
<td>0.91</td>
<td>0.78</td>
</tr>
<tr>
<td>BAI</td>
<td>0.87</td>
<td>0.87</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>PSS</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

\(a\) AUC: area under the curve.

\(b\) PHQ-9: Patient Health Questionnaire-9.

\(c\) BAI: Beck Anxiety Inventory.

\(d\) PSS: Perceived Stress Scale.

### Objective 2: Importance of Exposure to Mindfulness Videos in Comparison With Sociodemographics and Self-Reported Features in Predicting Intervention Success

After the introduction of the total exposure to the mindfulness videos to the data set, new predictive models were built (Table 5).
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Logistic regression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9(^b)</td>
<td>0.62</td>
<td>0.63</td>
<td>0.57</td>
<td>0.67</td>
</tr>
<tr>
<td>BAI(^c)</td>
<td>0.60</td>
<td>0.60</td>
<td>0.61</td>
<td>0.59</td>
</tr>
<tr>
<td>PSS(^d)</td>
<td>0.79</td>
<td>0.80</td>
<td>0.85</td>
<td>0.74</td>
</tr>
<tr>
<td><strong>Support vector machine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.78</td>
<td>0.75</td>
<td>0.96</td>
<td>0.61</td>
</tr>
<tr>
<td>BAI</td>
<td>0.93</td>
<td>0.93</td>
<td>0.86</td>
<td>1.00</td>
</tr>
<tr>
<td>PSS</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Random forest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.84</td>
<td>0.83</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>BAI</td>
<td>0.90</td>
<td>0.90</td>
<td>0.86</td>
<td>0.94</td>
</tr>
<tr>
<td>PSS</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Decision tree</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.84</td>
<td>0.81</td>
<td>0.96</td>
<td>0.72</td>
</tr>
<tr>
<td>BAI</td>
<td>0.80</td>
<td>0.80</td>
<td>0.86</td>
<td>0.75</td>
</tr>
<tr>
<td>PSS</td>
<td>0.97</td>
<td>0.97</td>
<td>1.00</td>
<td>0.94</td>
</tr>
<tr>
<td><strong>K-nearest neighbor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.81</td>
<td>0.78</td>
<td>0.96</td>
<td>0.67</td>
</tr>
<tr>
<td>BAI</td>
<td>0.84</td>
<td>0.83</td>
<td>0.89</td>
<td>0.78</td>
</tr>
<tr>
<td>PSS</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>Adaptive boosting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.84</td>
<td>0.81</td>
<td>0.96</td>
<td>0.72</td>
</tr>
<tr>
<td>BAI</td>
<td>0.93</td>
<td>0.93</td>
<td>0.89</td>
<td>0.97</td>
</tr>
<tr>
<td>PSS</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Gradient boosting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9</td>
<td>0.89</td>
<td>0.88</td>
<td>0.96</td>
<td>0.83</td>
</tr>
<tr>
<td>BAI</td>
<td>0.91</td>
<td>0.92</td>
<td>0.86</td>
<td>0.97</td>
</tr>
<tr>
<td>PSS</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

\(^a\)AUC: area under the curve.  
\(^b\)PHQ-9: Patient Health Questionnaire-9.  
\(^c\)BAI: Beck Anxiety Inventory.  
\(^d\)PSS: Perceived Stress Scale.

**Discussion**

**Principal Results**

The study investigated the predictability of the effectiveness of an MVC designed for undergraduate students to reduce symptoms of depression, anxiety, and stress as measured by PHQ-9, BAI, and PSS. The effectiveness was measured by the MCID for PHQ-9, BAI, and SPSS. Several algorithms were used to predict the MCID.

**Predicting Intervention Success With Sociodemographic and Self-Reported Measures**

We successfully built ML-based models that predicted the effectiveness of the MVC intervention. The highest AUC was achieved for gradient boosting to predict the intervention effectiveness for PHQ-9 and PSS (AUC=0.85 and AUC=1, respectively), followed closely by DT (AUC=0.84 and AUC=1, respectively) and AdaBoost (AUC=0.82 and AUC=0.99, respectively). The RF model had the highest AUC to predict BAI (AUC=0.93), followed closely by AdaBoost (AUC=0.92). AdaBoost might be the algorithm of choice for the 3 outcomes, as it is fairing a close second best for BAI and a close third best
for PHQ-9 and PSS. Gradient boosting and AdaBoost are both good choices to predict the intervention success for the 3 outcomes. It might be argued that AdaBoost might be preferable, given that it is usually less prone to overfitting than gradient boosting; however, there is no need to use the same algorithm to build the 3 predictors for the 3 outcomes.

We could not make a direct comparison with other studies that measured the 3 outcomes among university students using the same validated scales (PHQ-9, BAI, and SPSS). However, for PHQ-9, the performance of our model is higher than the one found in a previous study among adults in Korea using the Center for Epidemiologic Studies—Depression Scale 11 (AUC=0.87 and accuracy=0.86) [40] as well as the one found in a study in the United States that defined the success of the intervention as a 5-point reduction in PHQ-9 or a 4-point reduction in the General Anxiety Disorder screener-7 values (AUC=0.60 and accuracy=0.71) [35]. Regarding anxiety, the predictive model developed in this study had a higher performance (accuracy=0.92) than another study that used the Self-Rating Anxiety Scale, which did not report AUC but reported an accuracy of 0.84.

**Feature Importance**

Exposure to mindfulness videos was the most important factor in predicting the intervention’s success. This study has demonstrated a link between the MVC intervention’s success and exposure to mindfulness videos. It also confirms the results of the previous MVC pilot study that proved that exposure to mindfulness videos alone, without interaction between participants via an online discussion forum and without weekly videoconferencing with a coach, effectively reduced symptoms of depression, anxiety, and stress [26]. In other words, it indicates the ability of MVC to be deployed at a large scale without an increase in human resources. Scalability is a critical factor for eHealth intervention deployment in large populations. This finding suggests that scaling up an effective e-mental health MVC is possible in a cost-effective manner; scalability is one of the recognized failures in eHealth implementations [42].

**Practical and Policy Implications**

The MVC intervention does not provide clinical support; it is a platform that offers self-management of mental health symptoms (depression, anxiety, and stress). The MVC intervention proved to be effective [26-28] in reducing symptoms of depression, anxiety, and stress in university students. This study builds a predictive model that predicts intervention success using sociodemographic and self-reported measures; this will allow counseling services on university campuses to assess the usefulness of MVC for a particular student before taking the intervention and advise them accordingly to use MVC or to opt for another type of intervention. This will enable counseling services to personalize the advice to students’ profiles and allow students to manage their symptoms with the most appropriate intervention.

The other finding related to videos being the most important factor in predicting intervention success confirms the ability of MVC to be deployed at a large scale without an increase in human resources. The number of working hours is another important predictor of the success of the intervention. Although the provincial governments in Canada support university education, students must pay for their education and bear the cost of living. Not surprisingly, they work long hours, especially if they belong to a marginalized community. Our findings align with other studies that suggest that longer working hours outside the university and difficulty paying bills were recognized as predictors of poor mental health among students [43]. In Ontario, where the sample was taken, Statistics Canada recently reported an increased reliance of academic institutions on students’ fees in higher education to the extent that 54% of all college revenues in 2019/2020 were downloaded on students, which translates into an overall decline in public funding [44]. This situation pushed students to longer working hours; one can argue that since student debt has been recognized as negatively associated with mental well-being and academic outcomes [45,46], providing access to free higher education, supported by taxes such as in most of Europe, could enhance students’ mental well-being as it would relieve them from the need for long working hours.

**Strengths and Limitations**

One of the strengths of this study is the ability to predict the intervention’s success based on a few demographics and one question about self-rated health. Hence, the predictive model can be used in real life to indicate the suitability of online mindfulness intervention for specific individuals and possibly suggest alternatives if the model predicts noneffectiveness. The excellent AUC and accuracy measures make the models suitable for implementation and evaluation in real-life scenarios. However, the ML models must be monitored continuously if implemented for daily use (eg, a counseling service) [47,48].

A limitation of this study is that it relied on research done on 1 site; future research with larger samples with participants from multiple universities and colleges would better test the generalizability of results as it allows us to test the effectiveness of the models on external data.

**Conclusions**

Our results suggest that we can build high-performing models to predict MVC intervention effectiveness for depression, anxiety, and stress based on simple sociodemographics and self-reported features and that exposure to mindfulness videos is the most important predictor for the effectiveness of the intervention. Our findings provide evidence that scaling MVC can be done without additional cost for support and that the predictive models might be useful for professionals to advise students early enough on taking the intervention or choosing other alternatives.
Acknowledgments
The authors acknowledge the contribution of all students who spent valuable time participating in the study. The MVC Team members are Sahir Abbas, BSc; Yvonne Bohr, PhD; Manuela Ferrari, PhD; Wai Lun Alan Fung, MD, ScD, FRCP; Louise Hartley, PhD; Amin Mawani, PhD; Kwame McKenzie, MD, FRCP; and Jan E Odai, BA. They made contributions to several aspects of the project and the results’ development. They approve the final version and agree to be accountable for all aspects of the submitted paper. The work reported in this paper was funded by the Canadian Institutes for Health Research and eHealth Innovations Partnership Program grant (EH1-143553). The project’s principal investigators are CE, FA, and PR.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
CE, FA, and PR designed the original mindfulness virtual community study questionnaire, received the funds, and contributed equally. CE supervised FT, who performed and reported the analysis. CE verified the analysis and prepared the first draft. All authors provided critical feedback and revised it.

Conflicts of Interest
It is the understanding of the university and researchers that the Project Intellectual Property belongs to the CE, FA, and PR. The industry partner ForaHealthyMe.com owns all rights and titles to the copyrights of any computer source code software developed from this research project.

References


Abbreviations

AdaBoost: adaptive boosting
AUC: area under the curve
BAI: Beck Anxiety Inventory
DT: decision tree
KNN: k-nearest neighbor
LR: logistic regression
MCID: minimal clinically important difference
ML: machine learning
MVC: mindfulness virtual community
PHQ-9: Patient Health Questionnaire-9
PSS: Perceived Stress Scale
RCT: randomized controlled trial
RF: random forest
SVM: support vector machine
VC: virtual community

Please cite as:
El Morr C, Tavangar F, Ahmad F, Ritvo P, MVC Team
Predicting the Effectiveness of a Mindfulness Virtual Community Intervention for University Students: Machine Learning Model
URL: https://www.i-jmr.org/2024/1/e50982
doi:10.2196/50982
PMID:38578872
Designing mHealth Apps to Incorporate Evidence-Based Techniques for Prolonging User Engagement

Rebecca Monachelli¹, BS; Sharon Watkins Davis¹, MPA; Allison Barnard¹, BA; Michelle Longmire¹, MD; John P Docherty², MD; Ingrid Oakley-Girvan¹,³, MPH, PhD

¹Medable Inc, Palo Alto, CA, United States
²Weill Cornell Medical College, White Plains, NY, United States
³The Public Health Institute, Oakland, CA, United States

Corresponding Author:
Ingrid Oakley-Girvan, MPH, PhD
Medable Inc
525 University Ave
Palo Alto, CA, 94301
United States
Phone: 1 8778206259
Email: oakley@stanford.edu

Abstract

Maintaining user engagement with mobile health (mHealth) apps can be a challenge. Previously, we developed a conceptual model to optimize patient engagement in mHealth apps by incorporating multiple evidence-based methods, including increasing health literacy, enhancing technical competence, and improving feelings about participation in clinical trials. This viewpoint aims to report on a series of exploratory mini-experiments demonstrating the feasibility of testing our previously published engagement conceptual model. We collected data from 6 participants using an app that showed a series of educational videos and obtained additional data via questionnaires to illustrate and pilot the approach. The videos addressed 3 elements shown to relate to engagement in health care app use: increasing health literacy, enhancing technical competence, and improving positive feelings about participation in clinical trials. We measured changes in participants’ knowledge and feelings, collected feedback on the videos and content, made revisions based on this feedback, and conducted participant reassessments. The findings support the feasibility of an iterative approach to creating and refining engagement enhancements in mHealth apps. Systematically identifying the key evidence-based elements intended to be included in an app’s design and then systematically testing the implantation of each element separately until a satisfactory level of positive impact is achieved is feasible and should be incorporated into standard app design. While mHealth apps have shown promise, participants are more likely to drop out than to be retained. This viewpoint highlights the potential for mHealth researchers to test and refine mHealth apps using approaches to better engage users.

(KEYWORDS: adherence; app design; attrition; mHealth; user engagement; user experience; proof-of-concept)

Introduction

Smartphones have a global penetration estimated at 3.9 billion users [1], enabling mobile health (mHealth) apps to reach even low-resource areas and underserved populations [2,3]. mHealth apps have been developed to enable remote participation in clinical trials [4-7] and provide health education, health management, and other uses across the continuum from prevention through active treatment to palliative care [8]. Decentralized clinical trials using mHealth technologies promise faster participant accrual and a higher return on investment than traditional site-based trials [9]. mHealth apps have been shown to reduce inpatient readmission rates and decrease the length of hospital stay [10]. mHealth can increase knowledge and improve confidence and communication with health professionals [11]. However, while participants readily sign up for mHealth education and decentralized clinical trial apps, retention remains a major challenge [12-14]. For mHealth apps to succeed, users must consistently engage with them [15,16].

Engagement in mHealth apps has been conceptualized to include behavior, cognition, and affective components [17]. However, measures of patient engagement are underreported and lack consistency [18,19]. Participants are more likely to drop out than be retained despite app elements such as feedback,
reminders, in-app support, gamification, and participant compensation [20-23].

We developed a conceptual model to optimize patient engagement based on different phases of the engagement process [24]. Because digital literacy and anxiety have been shown to be negatively correlated with engagement [25], we established an approach to develop and test the educational components of our conceptual model to enhance app engagement by increasing health literacy, enhancing technical competence, and improving feelings about clinical trials. This Viewpoint aims to report on a series of exploratory mini-experiments, demonstrating the feasibility of testing our engagement conceptual model.

**How We Conducted the Exploratory Mini-Experiments**

**Testing Design**

We used a product testing approach rather than the traditional research evaluation approach. We used a group of existing product testers who are patients or caregivers working for Medable to rapidly test different iterations of our educational videos. Questionnaires were used both before and after participants viewed the videos, and semistructured interviews were also conducted.

**Data Collection**

We developed apps to collect specific data from participants over 1 week’s duration through questionnaires available on their smartphones before and after exposure to videos, as shown in Table 1. The videos were based on a review of the literature defining and studying each of these 3 target areas: health literacy, technical competence, and feelings about participation in clinical trials [24]. Each concept area was tested separately, with questionnaires specific to the educational component.

### Table 1. Schedule of tasks or questionnaires.

<table>
<thead>
<tr>
<th>Task or questionnaire</th>
<th>Frequency</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sign up</td>
<td>Once</td>
<td>Task 1</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Demographics (age, gender, education, housing, race, ethnicity, location, health condition, and notification)</td>
<td>Once</td>
<td>Task 2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Technology competence questions (TAM&lt;sup&gt;b&lt;/sup&gt;&lt;sup&gt;, CANX&lt;sup&gt;c&lt;/sup&gt;,&lt;/sup&gt; worries about pressing the wrong button and device preference)</td>
<td>Twice</td>
<td>Task 3&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Task 2&lt;sup&gt;e&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Technology competence video combined with practice questions (video on how to answer questions and practice questions)</td>
<td>Once</td>
<td>N/A</td>
<td>Task 1&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Health literacy questions (BRIEF&lt;sup&gt;f&lt;/sup&gt;, own health knowledge, clinical trial knowledge, and BMI)</td>
<td>Twice</td>
<td>N/A</td>
<td>N/A</td>
<td>Task 1&lt;sup&gt;d,ed&lt;/sup&gt;</td>
<td>Task 2&lt;sup&gt;e&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Health literacy video with knowledge check questions (St. Luke’s University Health Network’s video, “Wellness 101 – How to Improve Your Overall Health”; knowledge check questions)</td>
<td>Once</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Task 1&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Clinical trials question (temperature scale regarding participant’s feelings about study participation, Figure S1 in Multimedia Appendix 1)</td>
<td>Twice</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Task 1&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>Task 2</td>
</tr>
<tr>
<td>Clinical trials video (National Institute of Diabetes and Digestive and Kidney Diseases video: “Why Should I Join a Clinical Trial?”)</td>
<td>Once</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Task 1&lt;sup&gt;d,e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Study complete</td>
<td>Once</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Task 3</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.

<sup>b</sup>TAM: Technology Acceptance Model 3.

<sup>c</sup>CANX: Computer Anxiety.

<sup>d</sup>Set up with an automated morning reminder: “You have new tasks available today in the Patient Engagement app!”

<sup>e</sup>Set-up with an automated 8 pm reminder: “You have uncompleted tasks in the Patient Engagement app. Please finish them before midnight!”

<sup>f</sup>BRIEF: Brief Health Literacy Screening Tool.
In addition to collecting data through the questionnaires noted in Table 1, we conducted individual semistructured interviews with each participant at the end of the series of user evaluations using a video conferencing platform. Based on feedback, we revised specific videos and content and then conducted a second round of feedback. In this second round, after all the questions were answered in each section, participants were asked for immediate feedback with the open-ended question, “What did you think of this video?”

Recruitment
We used a product testing approach rather than a research evaluation approach. We recruited 6 individuals from the Medable Patient Care Network (PCN) who participated in this product development effort from May 2022 through July 2022. The PCN is a group of patients and caregivers who provide insights and user feedback from their perspective for a variety of apps being developed as products at Medable.

Ethical Considerations
This work was conducted and approved under the the Advarra IRB (Pro00062352). PCN members were paid an hourly rate of approximately US $150 by Medable for their work on behalf of the network in support of Medable product development efforts. Informed consent was obtained from all participants. All data were deidentified.

Assessment and Interventions

Health Literacy
The Brief Health Literacy Screening Tool [26] was selected to measure change before and after the health literacy intervention video. This questionnaire has four items that are rated on a 5-point Likert scale from “always” to “never”: (1) How often do you have someone help you read hospital materials? (2) How often do you have problems learning about your medical condition because of difficulty understanding written information? (3) How often do you have a problem understanding what is told to you about your medical condition? (4) How confident are you filling out medical forms by yourself? Two questions were asked in addition to the Brief Health Literacy Screening Tool using a 10-point scale: “How much do you know about your own health,” “How much do you know about clinical trials,” and 1 true or false question: “Do you know your body mass index (BMI)?” (Multimedia Appendix 1). Before showing a video, a module with a knowledge check portion was included to facilitate pre- and postvideo knowledge comparison. St. Luke’s University Health Network’s video, titled “Wellness 101 – How to Improve Your Overall Health” [27] was chosen for the content area of health literacy. The video provided 5 tips to improve an individual’s overall health. Five knowledge-related questions based on the video were asked. A BMI calculator was included as the final task in this section. Participants were instructed to “Try calculating your BMI on this website” using the CDC BMI calculator [28].

Technology Competence
We measured internet skills using a section of the Technology Acceptance Model 3 [29]. Statements presented to the participant included “The study website does not scare me at all,” “Working with the study website makes me nervous,” “The study website makes me feel uncomfortable,” and “The study website makes me feel uneasy.” We also asked, “While using the study website, I’m worried that I might press the wrong button and make a mistake that crashes the program” and “I am most comfortable using my (multi-select) iPad/Tablet, Smart Phone (iPhone or Android), Computer, Other.”

The video we used to increase technology competence was created in-house and showed participants examples of how to click on checkboxes, “radio button” response buttons, or move the cursor to a particular spot to answer different types of questions, including multiple choice, multiple selections, and a sliding scale. Participants were then asked to practice answering the same type of questions on their own (Multimedia Appendix 1 for the Technology Competence Questionnaire).

Clinical Trials
We also asked the question, “When it comes to your feelings about participating in this study, how do you rate your comfort?” using a scale ranging from “0, meaning no distress; totally relaxed” to “100, reflecting the highest anxiety/distress that you have ever felt” (Figure S1 in Multimedia Appendix 1). We showed the video from the National Institute of Diabetes and Digestive and Kidney Diseases, “Why Should I Join a Clinical Trial?” [30].

Semistructured Interviews
RM and AB conducted 45-minute interviews using Zoom (Zoom Technologies Inc) video conferencing with all 6 study participants using semistructured guides. Open-ended questions explored how participants felt about the process of using the app and how they felt about the questions that were asked through the app. A 5-point Likert scale was used to determine whether they agreed or disagreed with several statements focusing on how useful and informational they felt each of the videos and questionnaires were, for example, “I found the knowledge check questions to be useful” and “I felt less anxious about the idea of participating in a clinical trial after completing the knowledge check.”

Data Analysis
Given the small sample size and our product testing approach, we used simple descriptive statistics to give us insights into the differences between the “before” and “after” questionnaire results. The semistructured interviews were reviewed for commonalities.

What We Found

User Demographics
Users testing the smartphone apps ranged in age from 54 to 69 years. Of the 6 participants, 3 identified as female and 3 as male. Five stated their race as White, 1 selected Black or African American, and none identified as Hispanic. Participants lived in the United States and Europe; 5 owned their homes and 1 rented. The majority (n=5) indicated they had 1 or more health conditions, and 5 had completed at least 4 years of college. All participants indicated they preferred to receive notifications
before noon, 4 preferred SMS text message notifications, and 2 preferred email notifications.

**Health Literacy**

The mean score and SD for each survey item before and after viewing the instructional video are listed in Table S1 in Multimedia Appendix 1. Lower scores indicated a more positive response for 2 of the questions, specifically, “How confident are you filling out medical forms by yourself?” and “Do you know your Body Mass Index (BMI)?” The mean score for each survey item was slightly more positive after viewing the instructional video, except for the item “How often do you have a problem understanding what is told to you about your medical condition?” However, the SD was greater than the change in mean scores, indicating that it could be due to chance. There were no changes in questionnaire responses before and after watching the instructional video for 3 of the 6 participants, improvement in 2 questions and a decline in 1 question for 1 participant, improvement in 4 questions and no change in 3 questions for 1 participant, and improvement in knowledge of their overall health and a decline in the clinical trial knowledge item for 1 participant.

Participants’ feedback from the semistructured interviews revealed negative feelings toward the video “Wellness 101 – How to Improve Your Overall Health.” One participant described the video as “juvenile,” while another noted concerns that some participants might object to the medical video if they already smoke or have a high BMI. All participants agreed or strongly agreed to the question, “I found the knowledge check questions to be useful.” In total, 4 of the 6 participants neither agreed nor disagreed with the statement, “I felt less anxious about the idea of participating in a clinical trial after completing the knowledge check.” Participants liked the alternative health literacy video, “5 Ways to Make the Most of Your Doctor Visit” [31].

**Technology Competence**

The mean scores for all items were more positive after viewing the video. The scores for 3 participants improved after watching the instructional video but declined for 2 participants. There was no change for 1 participant (Table S1 in Multimedia Appendix 1).

Participants (N=6) shared positive feedback about the video showing how to answer and practice questions. When asked, “Watching somebody else demonstrate how to answer questions made me feel like I knew what was expected of me in the study,” participants answered, “agree” (2 participants) and “strongly agree” (4 participants). Additionally, participants answered “strongly agree” (4 participants) to the survey, “Practicing answering questions on my own made me feel less anxious about participating in the study.” Some participants felt the questions in this section were redundant and thought all the questions could be combined into 1 question. In addition, participants thought the questions felt negative with the emphasis on the terms “anxious” and “nervous” and suggested changing the questions to make them seem more positive. One participant suggested making the technology anxiety section optional for those who feel more comfortable using the study website.

**Clinical Trials**

Mean scores were slightly higher prior to viewing the instructional video compared with after the video. One participant improved by 10 points, 1 decreased by 10 points, and the other 4 stayed the same.

Participants had positive things to say about the video “Why Should I Join a Clinical Trial?” [30] When asked, “I found the video to be useful,” 2 participants answered “agree” and 4 answered “strongly agree.”

**Additional Overall User Feedback**

The participants had several specific suggestions. One participant suggested making sure that the videos were clearly specific to diseases or therapy areas in the trial and gave specific information on the trial structure. Two participants suggested including additional content on participant safety. We did an ad hoc assessment of this suggestion and sent the National Cancer Institute’s “Patient Safety in Clinical Trials” video for feedback [32]. The majority (n=5) participants liked the video. The National Institute of Mental Health’s video, “What are the risks and benefits of participating in clinical research?” was also considered [33]. Most (n=4) participants liked the video and 2 did not. Some participants thought the video was juvenile and better for children or young adults. Others also thought the video did not explain concepts such as placebos well. In response to this feedback, the following resources from the National Institute on Aging were added: “What are Clinical Trials and Studies?” [34] and “Clinical Research Benefits, Risks and Safety” [35], after which we sought a second round of feedback. Most participants (n=5) liked the additional resources.

**Discussion**

**Key Lessons**

Maintaining continuous and complete use of mHealth apps has remained a persistent problem that has not yielded even sophisticated solutions such as timed and individualized user messaging. A newer and evolving understanding of the foundational importance of user engagement with mHealth suggests that this problem comes from a lack of appreciation by mHealth app designers of the complex and multicomponent structures behind user engagement. We have built on prior knowledge and work to develop a model of engagement that accounts for the complexity of engagement [24].

This viewpoint was an exploratory study to determine the feasibility of this approach and to guide the refinement of this interactive test strategy. We learned several key lessons: (1) Specificity—the participants endorsed the recommendation that the interventions should be specific to the educational needs of the target of the mHealth app. The most positive feedback was given to the video we developed de novo to teach participants the technical competence required to correctly and effectively use the app to report their evaluations. (2) Attention to inadvertent adverse affective variables—the participants noted the importance of avoiding or rephrasing medical terms that...
could be seen as demeaning by some participants (eg, obesity, age, and infirmity). (3) Individualization—the participants clearly reflected different levels of need for improving their technical competence and health literacy. Our results indicate the potential importance of personalization in health app design addressing individuals’ levels of need and cultural and personal sensitivities. For example, a way to allow more individualization is to allow users to potentially opt out of certain learning features if they do not think they need them.

Comparison With Prior Work
Other studies have assessed how to adjust apps to increase engagement. One study found positive effects on adherence from personalization or tailoring of the app content to users’ needs, push notification reminders, user-friendly design, and personal support along with digital intervention [36]. However, the high dropout rate in app usage remains a major challenge [37]. A recent literature review found that despite factors such as appropriate reminders and feedback, app participants were more likely to drop out than be retained [20]. App literacy skills have been identified as a major factor in the uptake and engagement of smartphone apps [38]. Although we identified studies recommending web-based interventions to increase health literacy and technical skills [39,40], we have not found other studies testing approaches to increase those skills.

Limitations and Strengths
Our sample size was limited to 6 participants, 5 of whom were highly educated. We were unable to use any statistical significance measures because of the small sample size or draw conclusions that would apply to a larger population. Our highly educated sample is a limitation because these individuals may have better digital literacy than the general population. The main goal of our series of mini-experiments was to assess the feasibility of our approach and see whether we could retain the interest of the participants and obtain useful feedback on the interventions, which was successful. This type of testing is intended to evaluate tailored iterations of the app after gathering rapid participant feedback, with the ultimate goal of developing an app that will engage users. The next phase of our work will be to undertake the systematic testing of each component of this model in a larger and more diverse sample. We will then be able to refine the interventional enhancements for those components for a broader population. Ultimately, the functional use of this approach requires much larger, population-specific samples.

Conclusions
To fulfill the promise of mHealth apps to improve health outcomes, apps need to be improved so they reduce participant attrition. Health care apps do not work for people who do not use them. To date, app feedback, notifications and reminders, in-app support, gamification, and participant compensation have not been consistently successful in eliminating participant dropout. This study highlights the potential to develop and refine mHealth apps using evidence-based interventions derived from a broad range of behavioral and social science to increase engagement as a way of improving participant retention. This viewpoint highlights the potential for mHealth researchers to test and refine mHealth apps using approaches to better engage users. The preliminary experience reported in this viewpoint supports the feasibility of this iterative approach to create and refine engagement interventional enhancements for each element of the multidimensional, multicomponent theory of the engagement process.

Acknowledgments
We thank members of the Medable Patient Care Network for their valuable input and critical insights. No artificial intelligence products were used to develop questionnaires, interventions, analysis, or write any of this manuscript. Research reported in this publication was partially supported by the National Cancer Institute of the National Institutes of Health under contract numbers HHSN261201700030C and HHSN261201800010C. 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
The data sets generated and analyzed during this study are not publicly available due to the data release requirements of the Patient Care Network and the intellectual property associated with this work but are available from the corresponding author on reasonable request.

Authors’ Contributions
RM, AB, JPD, and IO-G conceived the mini-experiments designed to test the model developed by JPD and IO-G. RM and AB developed the app to collect data, coordinated with the Patient Care Network, and collected and analyzed the data. RM, JPD, IO-G, and SWD wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest
RM, JPD, SWD, and AB were employed by Medable at the time the data collection and manuscript writing was done. ML and IOG are currently employed at Medable. Medable is a clinical trial software-as-a-service platform and evidence-generation company and supported the authors’ conduct of this work without interference.
Multimedia Appendix 1
Study questionnaires and responses.
[DOCX File, 158 KB - ijmrv13i1e51974_app1.docx]


31. 5 ways to make the most of your doctor visit. National Institute on Aging YouTube page. 2019. URL: https://www.youtube.com/watch?v=36Sd8WpgR94

32. What are clinical trials and studies? National Institute on Aging. URL: https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies#:~:text=Clinical%20trials%20are%20research%20studies,safe%20and%20effective%20in%20people


35. What are clinical trials and studies? National Institute on Aging. URL: https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies#:~:text=Clinical%20trials%20are%20research%20studies,safe%20and%20effective%20in%20people


Abbreviations

mHealth: mobile health
PCN: Patient Care Network
Viewpoint

Using Routine Data to Improve Lesbian, Gay, Bisexual, and Transgender Health

Catherine L Saunders
Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Catherine L Saunders, MBBS, PhD
Department of Psychiatry
University of Cambridge
Herchel Smith Building, Forvie Site
Robinson Way, Cambridge Biomedical Campus
Cambridge, CB2 0SZ
United Kingdom
Phone: 44 1223337106
Email: cs834@medschl.cam.ac.uk

Abstract

The collection of sexual orientation in routine data, generated either from contacts with health services or in infrastructure data resources designed and collected for policy and research, has improved substantially in the United Kingdom in the last decade. Inclusive measures of gender and transgender status are now also beginning to be collected. This viewpoint considers current data collections, and their strengths and limitations, including accessing data, sample size, measures of sexual orientation and gender, measures of health outcomes, and longitudinal follow-up. The available data are considered within both sociopolitical and biomedical models of health for individuals who are lesbian, gay, bisexual, transgender, queer, or of other identities including nonbinary (LGBTQ+). Although most individual data sets have some methodological limitations, when put together, there is now a real depth of routine data for LGBTQ+ health research. This paper aims to provide a framework for how these data can be used to improve health and health care outcomes. Four practical analysis approaches are introduced—descriptive epidemiology, risk prediction, intervention development, and impact evaluation—and are discussed as frameworks for translating data into research with the potential to improve health.

Introduction

Research into health for individuals who are lesbian, gay, bisexual, transgender, queer, or of other identities including nonbinary (LGBTQ+) has consistently found that these populations experience poorer outcomes [1], with particularly strong and consistent evidence around poorer mental health for lesbian, gay, and bisexual adults [2]. LGBTQ+ health research typically uses 2 broad frameworks, sociopolitical or psychosocial models (where experiences of discrimination, victimization, stigma, and harassment are central to our understanding of health), and models of clinical, biomedical, and lifestyle risk factors [3]. Sexual health and HIV research falls mostly within biomedical frameworks; in practice, inequalities are interconnected [4,5]. Experiences of discrimination, harassment, and stigma have a profound impact on clinical health outcomes for LGBTQ+ communities [6]. Poorer health care access and quality among vulnerable groups exacerbate these impacts [7,8].

Data-driven LGBTQ+ health research has historically been based on convenience or purposive, rather than population-based samples [9], even in large-scale studies such as the National LGBT Survey in 2018 [1]. The collection of data on sexual orientation, gender, and transgender status has not been prioritized in routine sources [3]. Challenges in identifying understandable, meaningful, and acceptable measures, and concerns about LGBTQ+ respondents being able to answer questions safely, have been additional barriers to data collection [10].

However, data are improving [11]. The Equality Act in 2010 placed a statutory duty on public bodies in the United Kingdom...
to publish equality outcomes and report on progress in addressing disadvantage experiences by (among other characteristics) sexual orientation and gender reassignment [12], which had a strong positive impact. The Office for National Statistics has carried out development work on measures [13], and although the census in 2011 did not include questions about sexual orientation and transgender status, in part because of the concerns about respondent safety [10], by 2021 questions on both were included, with instruction statements that the questions were voluntary. In general, data collection for sexual orientation is more established than collection of data on transgender status and gender, although, for both, collections are improving.

This paper aims to provide a guide for how these improvements in routine data collection can potentially translate to improved health and health care outcomes.

Pathways From Data to Improved Health

LGBTQ+ health research using routine data sits within wider health data science research frameworks which are designed to leverage person-level routine health data (collected either from routine contacts with services or from infrastructure data resources designed and collected for policy and research) to improve outcomes [14]. Analyses fall under four broad translational pathways: (1) descriptive epidemiology, (2) risk prediction, (3) informing innovation and improvement, and (4) impact evaluation.

These pathways have the potential to improve outcomes by (1) providing evidence to inform policy and practice (descriptive epidemiology), (2) better targeting of interventions and understanding of population health needs (risk prediction), (3) more rational health service developments (intervention development), and (4) information on effectiveness informing commissioning or funding decisions (impact evaluation), respectively.

Health data science as a field has struggled with equality, diversity, and inclusion [15] and there are problems across the whole discipline. Algorithmic biases in risk prediction models, including in how models are developed, with differentially poorer functioning for minoritized groups, or inequitable outcomes when the models are implemented, are currently a particular area of concern [16,17]. In addition, missing data contribute not only to poorer risk model development, but to a lack of basic descriptive epidemiology, informed intervention development, or equalities impact evaluation. For LGBTQ+ health research using routine data, the pathways to improved outcomes are the same, as are the challenges of missing data [18].

Data

There are 5 groups or types of routine UK data sources—where information about sexual orientation and gender or transgender status are either well established or now starting to be recorded, beginning to address this lack of data. In the same way that LGBTQ+ health research balances both societal and biomedical models the data sources, which now include a collection of sexual orientation (more likely) or gender and transgender status (beginning to be introduced) reflect a balance of routine data from social and health sources. The five groups are (1) social science or societal data collections (including Understanding Society [19], birth cohort studies [20], educational cohort studies [21], and census data); (2) general and specific health surveys primarily designed to understand population health (including the Health Survey for England [22] and the National Survey of Sexual Attitudes and Lifestyles (NATSAL) [23]); (3) health services or patient surveys primarily designed for health service quality improvement (including the General Practice Patient Survey [24,25] and the Cancer Patient Experience Survey [26]); (4) health cohort studies (UK Biobank [27] and Our Future Health [28]); and (5) health records (including primary care research databases such as the Clinical Practice Research Datalink [29,30], and secondary and community services data sets, including the improving access to psychological therapies and mental health services data sets [31], and registry data, for example, cancer registry data [32]).

Put together, the data are starting to form a comprehensive collection but for each resource, there are strengths and limitations or challenges. For example, the Equality and Human Rights Commission was able to draw on quantitative evidence and data about sexual orientation and gender reassignment in work, education, and health in the State of the Nation report on equality and human rights in Britain published in November 2023 [33]. However, data access, sample size, measures of sexual orientation, gender, and health outcomes, and the ability to carry out longitudinal analysis and data quality vary across sources.

In terms of access to data, social science collections are primarily accessed without cost through the UK Data Archive; for sensitive fields, which often include sexual orientation or gender and transgender status, additional safeguards are in place. The UK Biobank and Our Future Health are 2 large biomedical data research cohorts accessed through trusted research environments (secure data hosting platforms) with relatively low but nonzero costs to researchers [28,34]. For all sources, access to data can require time and perseverance [11]. In terms of longitudinal follow-up, the UK Biobank is a mature cohort study, for which recruitment began in 2006 before sexual orientation and transgender status were routinely collected. Questions are included instead about sexual history, which provides some insight [34]. In contrast, Our Future Health for which recruitment began in 2022 has an inclusive gender question and questions about both sexual history and sexual orientation but only baseline data collection to date (recruitment is ongoing) [28].

Sample size is often a trade-off with detail. Understanding Society is a household panel survey designed to provide estimates about how life in the United Kingdom is changing and what stays the same over many years, with linked health and social data [19,35]. In common with other longitudinal and cohort data collections, the sample size is relatively small (about 40,000 people at baseline), compared, for example, to the General Practice Patient Survey which is a large cross-sectional survey designed to evaluate health care quality, which has a much larger sample size (about 700,000 responses) but with
much less nuanced health and particularly, social measures recorded.

Pooling data across sources is an approach to increasing sample size [36], and again resources are improving, estimates across an in-depth range of health outcomes from the Health Survey for England using data from across 7 years have recently been published [22].

Data from electronic health records (EHRs), or data routinely recorded as part of clinical or health care encounters offer both detailed health outcome data and large sample size. The challenge is often that for EHR data collections, in contrast with research data infrastructure resources, or survey-based health data collections, measurement of sexual orientation and gender or transgender status is less good. In part, this is because these are resources not designed for research but primarily collected as clinical documentation. Pilots have begun to improve recording, to support audit and quality improvement. However, given both historic discrimination experienced by the LGBTQ+ community based on sexual orientation recording in medical records, and the interpersonal interlinking of recording and coming out to a care provider [37,38], this solution to data improvement is not simply a neutral, administrative process. The reluctance of health care providers to ask about sexual orientation is a second barrier [39]. Recent research using EHR has provided insight by looking at transgender patients in primary care records based on prescribing and clinical codes, and this is an exciting area of progress [29,30]. These approaches have their own challenges, however, with historic clinical codes including outdated and discriminatory terminology still present in some older coded EHR records [18]. Legal barriers to identifying transgender patients after transition provide a further barrier to research using EHR; legislative changes have been required for recent quantitative analyses [40].

There are notable areas where data are poorer. Data governance and ethical challenges mean that data collections are much less likely to collect information on sexual orientation or gender from children. For example, some research studies have used proxies or less detailed response options where exact measures of ethnicity, gender identity, or disability cannot be asked [41]. Although HIV and sexual health research are well-studied topics in LGBTQ+ health research overall [42], routine data are usually more strongly safeguarded and less available for research, although measured in some collections.

**Applied Methodology**

The improvement in data collection for LGBTQ+ health research in the last decade mean that the applied methodological research around the use of these data is also developing. Questions about the longitudinal consistency of self-reported sexual orientation and history have been explored; changes are more frequently reported at younger ages [34,43,44]. For sexual orientation, missing data have reduced over time since the question has been routinely introduced in surveys [45]. Secular trends are also being better understood [44,45], meaning that age, period, and cohort effects in LGBTQ+ health research can begin to be untangled [24]. Differential item functioning for new questions among groups for whom English is a second language is a current area of concern for new gender questions, although this is unlikely to be a methodological issue specific to these particular items. The challenges of longitudinal consistency in question wording needing to be balanced against requirements for relevant and up-to-date survey items is again a methodological challenge not specifically limited to questions about sexual orientation and gender. New, nuanced, tools for understanding gender are beginning to be developed [46]; however, space constraints in surveys mean that often only single items are asked. While free text or more in-depth response options (or allowing multiple rather than single responses) are more inclusive [47,48], these nuanced data are often excluded from quantitative reporting. Data for people who identify as asexual are very limited, as are data for people with variations in sexual characteristics.

**How Have These Data Translated Into Applied Research?**

As LGBTQ+ routine data are improving, the insights that come from descriptive epidemiological LGBTQ+ health research are also developing. For example, historically, studies using routine data have been able to consider cancer risk factors such as, smoking and alcohol consumption [49-51], more easily than rarer cancer outcomes such as incidence. Limited sample size and poorer measurement of outcomes mean that earlier studies looking at cancer were cross-sectional and could only consider cancer prevalence without disaggregation by diagnosis [52]. Larger cross-sectional data sets have allowed disaggregation of diagnoses among lesbian, gay, and bisexual patients with cancer, identifying disparities primarily in HIV and human papillomavirus–associated cancers [26]. More recent work has for the first time in the United Kingdom been able to look at the impact of higher smoking prevalence identified in earlier studies on lung cancer incidence, using the UK Biobank resource [53], connecting both biomedical and sociopolitical frameworks; the LGBTQ+ community has historically been targeted by tobacco marketing.

Inequalities in LGBTQ+ mental health outcomes have also been well established through a series of studies and meta-analyses looking at cancer were cross-sectional and could only consider cancer prevalence without disaggregation by diagnosis [52]. Larger cross-sectional data sets have allowed disaggregation of diagnoses among lesbian, gay, and bisexual patients with cancer, identifying disparities primarily in HIV and human papillomavirus–associated cancers [26]. More recent work has for the first time in the United Kingdom been able to look at the impact of higher smoking prevalence identified in earlier studies on lung cancer incidence, using the UK Biobank resource [53], connecting both biomedical and sociopolitical frameworks; the LGBTQ+ community has historically been targeted by tobacco marketing.

Inequalities in LGBTQ+ mental health outcomes have also been well established through a series of studies and meta-analyses looking at cancer were cross-sectional and could only consider cancer prevalence without disaggregation by diagnosis [52]. Larger cross-sectional data sets have allowed disaggregation of diagnoses among lesbian, gay, and bisexual patients with cancer, identifying disparities primarily in HIV and human papillomavirus–associated cancers [26]. More recent work has for the first time in the United Kingdom been able to look at the impact of higher smoking prevalence identified in earlier studies on lung cancer incidence, using the UK Biobank resource [53], connecting both biomedical and sociopolitical frameworks; the LGBTQ+ community has historically been targeted by tobacco marketing.

Inequalities in LGBTQ+ mental health outcomes have also been well established through a series of studies and meta-analyses looking at cancer were cross-sectional and could only consider cancer prevalence without disaggregation by diagnosis [52]. Larger cross-sectional data sets have allowed disaggregation of diagnoses among lesbian, gay, and bisexual patients with cancer, identifying disparities primarily in HIV and human papillomavirus–associated cancers [26]. More recent work has for the first time in the United Kingdom been able to look at the impact of higher smoking prevalence identified in earlier studies on lung cancer incidence, using the UK Biobank resource [53], connecting both biomedical and sociopolitical frameworks; the LGBTQ+ community has historically been targeted by tobacco marketing.

Inequalities in LGBTQ+ mental health outcomes have also been well established through a series of studies and meta-analyses looking at cancer were cross-sectional and could only consider cancer prevalence without disaggregation by diagnosis [52]. Larger cross-sectional data sets have allowed disaggregation of diagnoses among lesbian, gay, and bisexual patients with cancer, identifying disparities primarily in HIV and human papillomavirus–associated cancers [26]. More recent work has for the first time in the United Kingdom been able to look at the impact of higher smoking prevalence identified in earlier studies on lung cancer incidence, using the UK Biobank resource [53], connecting both biomedical and sociopolitical frameworks; the LGBTQ+ community has historically been targeted by tobacco marketing.

Inequalities in LGBTQ+ mental health outcomes have also been well established through a series of studies and meta-analyses looking at cancer were cross-sectional and could only consider cancer prevalence without disaggregation by diagnosis [52]. Larger cross-sectional data sets have allowed disaggregation of diagnoses among lesbian, gay, and bisexual patients with cancer, identifying disparities primarily in HIV and human papillomavirus–associated cancers [26]. More recent work has for the first time in the United Kingdom been able to look at the impact of higher smoking prevalence identified in earlier studies on lung cancer incidence, using the UK Biobank resource [53], connecting both biomedical and sociopolitical frameworks; the LGBTQ+ community has historically been targeted by tobacco marketing.

Inequalities in LGBTQ+ mental health outcomes have also been well established through a series of studies and meta-analyses looking at cancer were cross-sectional and could only consider cancer prevalence without disaggregation by diagnosis [52]. Larger cross-sectional data sets have allowed disaggregation of diagnoses among lesbian, gay, and bisexual patients with cancer, identifying disparities primarily in HIV and human papillomavirus–associated cancers [26]. More recent work has for the first time in the United Kingdom been able to look at the impact of higher smoking prevalence identified in earlier studies on lung cancer incidence, using the UK Biobank resource [53], connecting both biomedical and sociopolitical frameworks; the LGBTQ+ community has historically been targeted by tobacco marketing.
are often easier to improve than more tangible health outcomes. However, although disparities in primary care access, communication, and satisfaction were measured routinely between 2011 and 2017, inequalities experienced by lesbian, gay, and bisexual adults persisted across the time period [58].

Impact evaluation is a second pathway, therefore, where routine data are beginning to be used to provide insight with the potential to change the care process and improve LGBTQ+ health outcomes. The collection of sexual orientation information in the Improving Access to Psychological Therapies data set has allowed inequalities evaluation of these services for lesbian, gay, and bisexual adults, finding that they were not as effective as for heterosexual service users [31]. In contrast, an inequalities evaluation of the introduction of telephone triage in general practices using the GP Patient Survey found that although there was variation between practices in outcomes, for different groups of patients within the same practice, including lesbian, gay, and bisexual adults, there was no evidence of differential impact on access to primary care [59]. The Millennium Cohort Study has been used to understand the differential impact of the COVID-19 pandemic on sexual minority groups [57], as has Understanding Society [60].

Where to Next—Routine Data Analysis?

Routine data for LGBTQ+ health research are much better today, in 2024, than they were even 10 years ago. Sexual orientation has now been collected in many sources for over a decade and more diverse and inclusive gender measures are being introduced, and are established in some collections. Of course, measurement needs to continue and is continuing to improve, and there are limitations and barriers; no data set alone is perfect.

However, across the spectrum of sources, there is a real depth of data now available and in terms of research, the data are good enough now to at least start thinking properly about how we can use these resources to improve LGBTQ+ health and tackle inequalities.

In terms of data development, of course linkage is 1 exciting potential future avenue, with the linked 2011 census and routine health care data in Scotland providing a possible model for future development. But in reality, using routine data for LGBTQ+ health research lies within the wider UK research landscape for using routine data overall. Here the Goldacre review probably shines some light on the direction of travel [61]. Access is becoming more cautious, and data are becoming more securely safeguarded, new frameworks and solutions are needed to ensure that access continues and barriers do not increase [62]. For sensitive fields such as sexual orientation, gender, and transgender status, this is particularly important, but it is likely that time, patience, and perseverance are going to continue to be required when working in this space. As a balance to concerns about the use of person-level data, tools sharing aggregate data such as the census resources from the United Kingdom’s Office for National Statistics [63], and the analysis tool for the GP Patient Survey remain important resources and provide real insight.

So, the question remains, how are we going to use these data to improve LGBTQ+ health? Although the data are better, the approaches have not changed and the methodological answers to the pathways from data to improved health remain the same. The four pathways are (1) descriptive epidemiology, (2) risk prediction, (3) informing innovation and improvement, and (4) impact evaluation.

Given the recentness of the data improvements and that data resources are still improving, there remains a real need for basic epidemiological descriptive work using these new data to answer questions and provide insight where simply the data have not been available before. More in-depth analyses, analyses considering longitudinal changes, and better measures of health and health outcomes, as well as sexual orientation and gender and transgender status, are all part of this. Frameworks for addressing health inequalities require researchers to go beyond simply describing known inequities [14], but for LGBTQ+ health there is still an evidence gap where descriptive epidemiology that focuses on areas where research could have an impact on policy has a place.

Maybe the results will be unsurprising, and research may show that inequalities have not disappeared as the data have improved, but the work is still important, and insight is still needed.

Risk prediction as a field has real challenges ahead to get to grips with equality, diversity, and inclusion, and this needs to include LGBTQ+ health. For transgender health specifically, there are some more questions to ask around risk model development; the exclusion of transgender adults from the development of some risk scores [47], and lack of clarity about how to implement scores based on binary gender or sex classifications are some specific issues to add to these [18]. Although methodological work is still needed to understand the best way to develop and implement risk scoring for transgender patients to avoid potentially both under- and overtreatment, the critical first step is to ensure that data used for model development do not exclude transgender populations before the research begins.

In terms of intervention development and audit, the improvement of data is important to ensure that evidence-based interventions are developed and part of wider thinking about how routine data can improve health and LGBTQ+ health in particular. Specific clinical data sets, such as cancer data collections or more in-depth surveys such as NATSAL will be particularly important in this domain. Much local evaluation of LGBTQ+ health interventions remains qualitative [64], and the evidence base for health equity audits to address inequalities remains poor [65].

The importance of including explicit inequality analyses in impact evaluations remains a key analysis strategy for improving health. Even when interventions are not LGBTQ+ specific, there may or may not be an inequitable impact. This kind of routine equalities impact work for LGBTQ+ and other groups is central to the drive the Equality Act has given to the improvement in data that we have seen, and needs to become a routine part of evaluative work.
**Where to Next—LGBTQ+ Health Research?**

The data are good enough now for routine data to play a substantive part in LGBTQ+ health research, and there are clear and realistic pathways for how this research can potentially improve health. This comes within the wider context of flourishing LGBTQ+ health research overall [66]. Health and health care are complex [67]. It is not a linear pathway from data to improved health outcomes; but good research can play a part.

There are particular challenges for health research with LGBTQ+ children and young people, where data are often less frequently collected and ethical and governance considerations are particularly important, and there is an identified need for more research [68,69]. The co-option of research findings into homophobic or transphobic narratives is a further difficult area, as are avoiding some of the blind spots around equality, diversity, and inclusion in routine data research that are beginning to be identified particularly in risk prediction work [15]. Good communication and cautious interpretations of findings are part of the solution, as are patient and public involvement, and the involvement of LGBTQ+ communities in identifying research priorities and in carrying out research [55]. Best practice guidance for LGBTQ+ health research [70], inclusive public involvement [71], and involvement in LGBTQ+ health research [72] provide some signposts for researchers.

**Conclusions**

Descriptive epidemiology, risk prediction, informing innovation and improvement, and impact evaluation are 4 practical pathways from data to improved health. Data for LGBTQ+ health research are now good enough and improving. We know that health inequalities exist, within both societal and biomedical frameworks. Research with strong public involvement, good clear communication, and stakeholder involvement is key, as in all research. Overall, this is a positive story for routine data. We are at the stage where the analysis of routine data can contribute to making real practical steps toward informing policy and practice, better targeting of interventions and understanding of population health needs, more rational health service developments, informing commissioning or funding decisions, and improving LGBTQ+ health.

**Acknowledgments**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Conflicts of Interest**

None declared.

**References**


32. Interact J Med Res 2024 | vol. 13 | e53311 | p.32 https://www.i-jmr.org/2024/1/e53311 (page number not for citation purposes)


42. Pakianathan M, Daley N, Hegazi A. Gay, bisexual, and other men who have sex with men: time to end the fixation with HIV. BMJ 2016;354:i4739. [ doi: 10.1136/bmj.i4739] [Medline: 27590481]


https://www.i-jmr.org/2024/1/e53311

Interact J Med Res 2024 | vol. 13 | e53311 | p.33

(page number not for citation purposes)
70. Ethical research: good practice guide to researching LGBT communities and issues. LGBT Foundation. URL: https://dxfj8rzpbpyrw.cloudfront.net/Files/1a884870-453a-429d-a213-399a4502472e/Ethics%2520Guide.pdf [accessed 2024-04-10]

Abbreviations

EHR: electronic health record  
LGBTQ+: lesbian, gay, bisexual, transgender, queer, or other identities including nonbinary  
NATSAL: National Survey of Sexual Attitudes and Lifestyles

[accessed 2024-04-10]
Ad Hoc Modifications to a High Dependency Psychiatric Unit for People With Dementia During the COVID-19 Period

Thanita Pilunthanakul1*, BSc, MA, MD; Giles Ming Yee Tan2*, MBBS, M Med

1MOH Holdings Pte Ltd, Singapore, Singapore
2Institute of Mental Health, Singapore, Singapore
*all authors contributed equally

Corresponding Author:
Thanita Pilunthanakul, BSc, MA, MD
MOH Holdings Pte Ltd
1 Maritime Square
Singapore, 099253
Singapore
Phone: 65 66220980
Email: thanita.pilunthanakul@mohh.com.sg

Abstract

The COVID-19 pandemic led to behavioral exacerbations in people with dementia. Increased hospitalizations and lack of bed availability in specialized dementia wards at a tertiary psychiatric hospital in Singapore resulted in lodging people with dementia in the High Dependency Psychiatric Unit (HDPCU). Customizations to create a dementia-friendly environment at the HDPCU included: (1) environmental modifications to facilitate orientation and engender familiarity; (2) person-centered care to promote attachment, inclusion, identity, occupation, and comfort; (3) risk management for delirium; and (4) training core competencies. Such practical solutions can also be implemented elsewhere to help overcome resource constraints and repurpose services to accommodate increasing populations of people living with dementia.

(KEYWORDS)
dementia; COVID-19; high dependency psychiatric unit; psychiatric intensive care unit; caregiver stress; SARS-CoV-2; psychiatric; psychiatry; mental health; health care system; Alzheimer; ward; care facility

Background

The spread of SARS-CoV-2 causing COVID-19 required significant reorganization of the Singapore health care system to support the rising number of COVID-19 cases and associated mental health issues [1,2]. Postponement of nonurgent clinic appointments due to reallocating hospital resources for pandemic control and patients’ fear of contracting COVID-19 led to significant delays in treatment and further increased the risk of psychiatric relapses [3]. Reduced physical and cognitive activity from the cessation of dementia daycare programs and limited social engagements due to restricted visits with friends and family further increased social isolation and behavioral exacerbations in people with dementia [4]. On top of this, unemployment, work-from-home policies, quarantine orders, and home-based learning put family members in closer contact with people with dementia. This struggle to cope with the increasing care demands of people with dementia has resulted in greater caregiver burnout and hospitalizations for respite care [5,6].

Due to the shortage of beds in the dementia wards at the Institute of Mental Health (IMH), a tertiary psychiatric care facility in Singapore, people with dementia are occasionally lodged in the High Dependency Psychiatric Care Unit (HDPCU) of the IMH, which was not initially designed for dementia-friendly interventions. In this viewpoint, we provide insight into easily implementable and practical creative solutions that have been used to accommodate people with dementia in the HDPCU during the COVID-19 pandemic, overcoming resource constraints and repurposing services in the face of changing needs using a patient-centered approach (summarized in Table 1).
Dementia wards have specific modifications to orient people due to separation from familiar settings, people, and routines. People with dementia often experience disorientating situations to manage the patients. A psychiatrist, a junior doctor, and the nursing and allied health care team are on site applying physical restraints and administering oral and intramuscular sedation if required. A nurse-to-patient ratio. Staff are specially trained in swift de-escalation to ensure safety and prevent violence, including partition areas to create “rooms” for different activities, place photographs of friends/family close to the patient’s bed, arrange regular communication between the patient and their family/friends via video calls.

Table 1. Summarized framework for customizing a care environment for people with dementia.

<table>
<thead>
<tr>
<th>Principles</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental modifications to facilitate orientation and engender familiarity</td>
<td>- Clear signs and signage&lt;br&gt; - Sizable and readable calendars and clocks&lt;br&gt; - Frequent reorientation by staff&lt;br&gt; - Nurse in areas allowing natural sunlight and views of greenery&lt;br&gt; - Simulate home-like surroundings, including partition areas to create “rooms” for different activities&lt;br&gt; - place photographs of friends/family close to the patient’s bed&lt;br&gt; - arrange regular communication between the patient and their family/friends via video calls</td>
</tr>
<tr>
<td>Person-centered care to promote attachment, inclusion, identity, occupation, and comfort</td>
<td>- Obtain a detailed personal history from family/friends regarding the patient’s preferences&lt;br&gt; - Surround the patient with items that affirm their personhood, including playing songs that bring comfort or voice recordings of family/friends in the ward&lt;br&gt; - addressing patients by their usual/preferred nickname&lt;br&gt; - if possible, allow patients to wear/have sentimental items close by&lt;br&gt; - Empower patients to exercise choice as much as possible, no matter how small the decisions may be&lt;br&gt; - Encourage patients to join together for meals and games&lt;br&gt; - Engage patients in meaningful and mentally stimulating activities&lt;br&gt; - Express comforting interactions and validate patients’ concerns&lt;br&gt; - Patiently answer repeated questions and allow relatively more time to perform tasks</td>
</tr>
<tr>
<td>Risk management (delirium)</td>
<td>- Actively take measures to prevent delirium, including restrict physical restraints to only when necessary, and even then, for the shortest duration required&lt;br&gt; - minimize medications that risk iatrogenic delirium&lt;br&gt; - Obtain a corroborative history regarding patients’ expression of discomfort to recognize signs of distress and address agitated behavior promptly</td>
</tr>
<tr>
<td>Core competencies (training if required) of staff</td>
<td>- Understand core concepts of and practice person-centered care&lt;br&gt; - Geriatric-specific care, including fall and choking risks, along with activities of daily living support&lt;br&gt; - De-escalation skills for agitated geriatric patients</td>
</tr>
</tbody>
</table>

Environmental Modifications of the HDPCU

The HDPCU is a specialized inpatient unit devised for patients with an acute psychiatric disorder linked to severe agitation or aggression, placing them at significant risk to themselves or others, leading to the requirement of close monitoring. The nursing counter is sandwiched between two locked gender-specific cubicles with 4 and 6 beds, respectively, and single bathrooms. The nurse has a full view of both cubicles and there are various discreetly placed security cameras. Items that could fuel self-harming or suicidal behaviors, such as wires for electronics, plastic bags, detergents, sharp pencils, and utensils, are strictly prohibited in the ward. There is a 2:1 nurse-to-patient ratio. Staff are specially trained in swift de-escalation to ensure safety and prevent violence, including applying physical restraints and administering oral and intramuscular sedation if required. A psychiatrist, a junior doctor, and the nursing and allied health care team are on site to manage the patients.

People with dementia often experience disorientating situations due to separation from familiar settings, people, and routines. Dementia wards have specific modifications to orient people with dementia, such as legible signage; large clocks; brighter lighting; and contrasting-colored walls, furniture, and utensils [7]. Renovating the HDPCU to suit such requirements was not immediately feasible. Hence, modifying the environment to have clear signs indicating the bathrooms; a sizeable hand-drawn daily calendar facing the bed indicating the date, day, month, and year; and verbal reorientation 3 times a day were implemented to facilitate orientation. People with dementia were also nursed opposite a readable digital clock and beside big windows that offered a view of greenery and allowed in natural sunlight.

Dementia wards engender familiarity by creating homey surroundings, including paintings hung along corridors and divided kitchen, bedroom, and living room spaces. Since safety is of utmost priority in the HDPCU, rules are often strict, and cubicles are designed to be relatively smaller than found in other wards along with an open layout for easy monitoring. These restrictions may cause people with dementia, particularly those who like to wander, to feel trapped and anxious. Given the nature of patients admitted to the HDPCU, the noisy and disruptive atmosphere can destabilize and frighten people with dementia. To create a calm environment that minimizes overstimulation and distractions, people with dementia were nursed in a partitioned visitors’ area accessed via a corridor.
adjacent to the cubicles and nursing counter. Families of people with dementia were encouraged to bring photographs to place in front of patients’ beds and participate in regular video calls from the ward smartphone to lessen the effects of visitor restrictions during the pandemic [8]. The sectioned area also reduced the risk of impaired sleep-wake cycles and sun-downing behaviors that are common among people with dementia and could provoke other patients. The improvised space simulated a private bedroom, while a wheel-in television and movable couches in the cubicles’ shared living area imitated a makeshift living room.

Promoting Person-Centered Care

In person-centered care for people with dementia, personhood consists of attachment, inclusion, identity, occupation, and comfort. Emotional distress is usually triggered by unmet needs related to aspects of personhood. Obtaining a detailed personal history from the family regarding the preferences of people with dementia is essential to affirm personhood. For one such patient, playing Chinese songs from his childhood, hearing voice recordings of family, addressing him by his preferred nickname, and wearing a jacket gifted from his daughter in the ward provided a sense of comfort, identity, and continuation of self. Empowering the patient to exercise choice as much as possible, even for tasks as small as choosing a preferred snack, preserves autonomy and dignity. Encouraging people with dementia to join other patients during meals and games instills a sense of inclusion and occupation [9,10]. Engaging people with dementia in meaningful activities mentally stimulates and reduces the restlessness related to the tendency to worry about their situation. Expressing warmth through comforting interactions, patiently answering repeated questions, allowing them more time to perform tasks, and validating concerns can help to settle the wariness and diminished sense of attachment experienced by these patients. These efforts promote the therapeutic relationship and trust between staff and people with dementia, ultimately reducing aggression and distress.

Risk Management

Lastly, because people with dementia are prone to delirium during acute hospitalization, the HDPCU team actively took measures to prevent this risk. Physical restraints were only applied if verbal de-escalation repeatedly failed and the extreme agitation posed a safety risk to themselves or others; when required, patients were restrained for the shortest duration necessary. Wherever possible, medications that risk iatrogenic delirium in people with dementia were avoided, such as short-acting benzodiazepines for tranquilization, anticholinergic drugs, and opioid-containing analgesics. People with dementia often have issues communicating their needs and are likely to only respond to their present state due to verbal difficulties and memory problems. Obtaining a further history regarding the patients’ typical behavioral patterns and expression of discomfort from pain, hunger, thirst, or constipation helped the team promptly recognize signs of distress and address agitated behaviors early without escalating to restraints.

Prospects

In conclusion, hospital care conditions can be difficult for people with dementia as they require familiarity, frequent orientation, and a high level of staff trained to handle their needs. The HDPCU adapted to rapid hospital protocol and health care policy changes during the COVID-19 pandemic and the resultant rise in the inpatient dementia population. Although the HDPCU staff were not geriatric-trained, the favorable staffing ratio and expertise in handling agitated and aggressive patients made it easier to implement person-centered care. Such conditions may not be available in nonspecialized wards, posing a challenge for catering to the increasing population of people with dementia admitted to hospitals in Singapore. Nonetheless, creative solutions could be established to customize the environment for such patients aptly. Hospitals could also consider bringing in key “experts” such as psychogeriatricians and geriatric nurses to advise on optimizing nonspecialized wards and provide training to care for people with dementia.

Acknowledgments

The authors thank the nurses and allied health care team from the High Dependency Psychiatric Unit at the Institute of Mental Health, Singapore, for their team effort and support.

Authors’ Contributions

TP contributed to conceptualizing and writing the original draft and to reviewing and editing the manuscript. GTMY contributed to conceptualizing, supervising, and reviewing and editing the manuscript.

Conflicts of Interest

None declared.

References


Review

Strategies to Alleviate the Burden Experienced by Informal Caregivers of Persons With Severe Mental Disorders in Low- and Middle-Income Countries: Scoping Review

Olindah Silaule¹, MSc; Daleen Castelein², PhD; Fasloen Adams³, PhD; Nokuthula Gloria Nkosi⁴, PhD

¹Department of Occupational Therapy, University of the Witwatersrand, Johannesburg, South Africa
²Department of Occupational Therapy, University of Pretoria, Pretoria, South Africa
³Division of Occupational Therapy, Stellenbosch University, Cape Town, South Africa
⁴Department of Nursing Education, University of the Witwatersrand, Johannesburg, South Africa

Corresponding Author:
Olindah Silaule, MSc
Department of Occupational Therapy
University of the Witwatersrand
7 York Road
Parktown
Johannesburg, 2193
South Africa
Phone: 27 0782035213
Email: osilaule@cartafrica.org

Abstract

Background: There is considerable evidence of the burden of care encountered by informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. Previous studies have highlighted the need to support these informal caregivers as key players in the care of these patients. To date, limited evidence exists on the extent and types of strategies for supporting these informal caregivers in low- and middle-income countries.

Objective: This scoping review aims to identify and describe the extent and type of evidence on the existing strategies for alleviating the burden of care among informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries.

Methods: A systematic literature search was completed following the Joanna Briggs Institute methodology for scoping reviews. The participants, concept, and context framework was used to guide the search for literature sources across 5 databases: PubMed, MEDLINE, CINAHL, and PsycINFO for published literature and ProQuest for unpublished literature. This review included studies that reported on strategies for alleviating the burden of care among informal caregivers of persons with severe and enduring mental health conditions, with a focus on studies that evaluated or recommended caregiver interventions and support strategies in low- and middle-income countries. The search was limited to studies conducted between 2001 and 2021, and only papers written in English were considered for inclusion. Using the Covidence software (Veritas Health Innovation), 2 reviewers independently screened the papers, applied the inclusion and exclusion criteria, and met biweekly to discuss and resolve conflicts. The relevant studies and reported outcomes were summarized, organized, and analyzed descriptively using numeric summary analysis and deductive content analysis.

Results: Of the 18,342 studies identified, 44 (0.24%) met the inclusion criteria. The included studies were from 16 low- and middle-income countries in Asia, Africa, Europe, and South and North America. Most studies (21/44, 48%) were randomized controlled trials conducted in Asian countries. The identified strategies were grouped into 2 categories: implemented and recommended intervention strategies. Identified strategies included community-based interventions, psychoeducation interventions, support groups, cognitive behavioral therapy, spirituality-based interventions, and smartphone-based interventions. In addition, mindfulness and empowerment, collaborative interventions, standard care, financial and social support, counseling, occupation-based interventions, policy and legislature, and access to mental health care were identified. Psychoeducation and support group interventions were identified as common strategies for alleviating the burden of care among informal caregivers of persons with severe and enduring mental health conditions.
Conclusions: This review provides evidence on the types of implemented and recommended strategies for alleviating the burden of care among informal caregivers in low- and middle-income countries. Although psychoeducational interventions were the most preferred strategy for alleviating burden, their benefits were short-lived when compared with peer-led support groups.

International Registered Report Identifier (IRRID): RR2-10.2196/44268


KEYWORDS
severe mental disorders; informal caregivers; caregiver stress; caregiver support; low- and middle-income country; mobile phone

Introduction

Background Functional Implications of Severe and Enduring Mental Health Conditions

People with severe and enduring mental health conditions are unable to cope with the demands of everyday life and need care after being discharged from a health facility. Serious functional limitations frequently interfere with the ability of people with severe and enduring mental health conditions to perform essential roles such as being a worker, family member, or friend [1]. The lack of independence among people with severe and enduring mental health conditions in daily living activities and, in some cases, behavioral problems result in poor quality of life. Psychosis, bipolar mood, major depressive, anxiety, eating, and personality disorders are often classified as severe and enduring mental health conditions [2]. The amount of care needed varies greatly depending on the diagnosis, severity of symptoms, and level of independent functioning. As such, assistance is needed in areas of basic self-care, healthy eating, following daily routines, medication management and compliance, engagement in meaningful activities, and community integration. Therefore, informal caregivers are essential to fulfill the care needs of people with severe and enduring mental health conditions. This task becomes draining, and the caregiver burden should not be underestimated [3,4].

Informal Caregivers in Mental Health

Informal caregivers are people who deliver care without remuneration to persons with a chronic condition. The caregivers of persons with severe and enduring mental health conditions are most often family members or neighbors who assist with the care of the person with a mental health condition [5]. Dixon et al [6] noted that, often, the caregiver of a person with a severe and enduring mental health condition does not have a choice in being the caregiver, which is an additional burden of care for the carer. Informal caregivers’ roles include monitoring medication, being the contact person between the health provider and family, early identification of signs of relapse, taking care of daily tasks such as self-care, providing meals, and ensuring the safety of the person [7,8]. Informal caregivers provide emotional support when needed and deal with the challenging behavior of the person with a severe and enduring mental health condition, which may lead to police involvement [9]. Involving informal caregivers in the routine care and management of persons with severe and enduring mental health conditions has shown a positive influence on the course of the illness but only if the caregivers’ needs are addressed and they are supported in one way or another [5,6,10].

Caregiver Burden and Need for Support

The burden of care experienced by caregivers is classified into objective and subjective burden. Objective burden refers to the tangible impact of the demands of caring tasks and encompasses the practical and concrete aspects of caregiving that can be quantified or assessed externally [11,12]. Subjective burden refers to the emotional and psychological experiences, feelings, and perceptions of caregivers related to their caregiving role [8,13]. It focuses on how caregivers perceive the impact of their responsibilities on their well-being and mental health.

In low- and middle-income countries, objective burden can easily overshadow subjective burden, such as a lack of community mental health care services and clinics that are out of stock of medication for people with severe and enduring mental health conditions, causing relapse and often readmission [11]. Other concrete aspects of objective burden, such as long hours of being available to the care recipient, time lost from daily activities, and not being able to earn an income, have serious consequences for caregivers as they are unable to pursue their own goals in life and, accordingly, experience a lower quality of life [9,12]. Cultural beliefs about being cursed by ancestors [14] and stigma from health care professionals [15] further aggravate the caring duties of caregivers, precipitating objective burden. The objective burden of care is less reported in the literature than subjective feelings of burden. Many studies have reported the subjective burden of care, which highlights emotional distress and feelings of anxiety, depression, and sadness related to the challenges of caregiving and witnessing the struggles of their family member. Loss of freedom and autonomy has also been reported owing to the demand of caregiving [16]. Feelings of shame and social isolation stem from the stigma of mental illness and cultural beliefs about being cursed by ancestors. Caregivers can also experience role strain as they have to juggle responsibilities at home, which can cause fatigue and overall dissatisfaction in life [17]. Both dimensions of burden are crucial for understanding the challenges faced by family caregivers of individuals with severe and enduring mental health conditions and play a vital role in informing support strategies and interventions.

There is overwhelming evidence in the literature for the need to support caregivers, and several studies from high-income and low- and middle-income countries have indicated successful strategies for alleviating caregiver burden. Studies on informal caregivers of people with severe and enduring mental health conditions in low- and middle-income countries have also increased over the past 10 years, and the burden of caregivers has been well described [18-21]. Yerriah et al [16] reported on the extent of the burden of caregivers of persons with severe mental disorders; informal caregivers; caregiver stress; caregiver support; low- and middle-income country; mobile phone
Support Strategies for Caregivers

To date, a number of strategies to support caregivers have been investigated with varying results. In Turkey, family-to-family support programs have shown a positive impact on the burden of care [22]. A meta-analysis by Chen et al [23] showed some support for nonpharmacological interventions (mostly psychoeducation) for caregivers of persons with schizophrenia, but the authors reported a potential bias in the results because of the small sample size. In contrast, the systematic review and meta-analysis by Sin et al [24] did not support psychoeducation to improve compliance with treatment and prevent relapse in persons with psychosis, and they reported a lack of available data; thus, no meta-regressions could be conducted. Ewertzon and Hanson [18] conducted a narrative review and identified provision of knowledge, problem-solving, stress management, mutual support groups, and individual-support interventions as successful support interventions for caregivers. Finally, a systematic review by Napa et al [25] revealed that there is insufficient evidence of interventions for psychological distress and expressed emotions in families of persons who experienced first-episode psychosis.

There is limited evidence of web-based health care services and digital health technologies for supporting informal caregivers of individuals with severe and enduring mental health conditions [26]. Sin et al [27] developed an eHealth intervention called Carers for People with Psychosis e-support, but its effectiveness has not yet been investigated. Ploeg et al [28] conducted a rapid review of web-based interventions to improve general caregiver outcomes. More than half of the 17 included studies showed a positive outcome for decreased depressive symptoms, stress, and anxiety among caregivers. If virtual strategies can support caregivers in rural areas, they may be a feasible method to reach people in remote areas who have poor access to health care.

Informal caregivers in rural or remote areas face additional objective burdens such as poor access to services, lack of integration of mental health into community health services [29], and poor intersectoral collaboration. They often have to wait long hours before any support arrives, and in many cases, they have to deal with challenges with the limited resources they have available.

If the types of intervention strategies that could be relevant for low- and middle-income countries, how they were implemented, and the outcome that was achieved were mapped, it could guide health care workers to support informal caregivers on various levels and with various strategies. The availability of virtual support strategies could add another dimension of support to caregivers, which may lead to additional positive outcomes such as immediate support, available information, and contact with support groups. People with severe and enduring mental health conditions benefit as they are most likely to receive optimal care, and their relapse rate may decrease, which means less need for readmission in overextended mental health care wards or hospitals. In addition, people with severe and enduring mental health conditions may also experience a better quality of life if their carers are supported. Thus, it is essential to understand the strategies to alleviate informal caregiver burden and how these strategies should be implemented.

This scoping review aimed to map the strategies to alleviate the objective and subjective burden of informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. The objectives of this scoping review were to (1) identify the types of existing strategies (virtual and face-to-face) for alleviating the objective or subjective burden of care, (2) describe the characteristics of the identified strategies, and (3) list the positive outcomes that were achieved using the identified strategies.

Methods

Review Methodology

This scoping review followed the Joanna Briggs Institute methodology specific to scoping reviews [30]. An a priori protocol for this review has been published [31].

Review Question

The scoping review addressed two research questions related to the strategies for alleviating the burden of care among informal caregivers of persons with severe and enduring mental health conditions:

1. Which existing strategies are reported in the literature for alleviating the burden of informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries?
2. What are the outcomes reported by the authors of the strategies for alleviating burden among informal caregivers?

Eligibility Criteria

Participants

Studies were included if participants were informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. Studies that focused on informal caregivers of patients diagnosed with Alzheimer disease and dementia were excluded. This included family, friends, neighbors, and community members who voluntarily provided care without any remuneration. Studies that included both informal caregivers and patients as participants were considered for inclusion, whereas those that reported only on patients were excluded.

Concept

Caregiver burden was defined as the physical, psychological, emotional, and financial stresses of providing care to a person with severe and enduring mental health conditions [32]. Severe and enduring mental health conditions in this scoping review included the schizophrenia spectrum and other psychotic disorders, bipolar and related disorders, mood or depressive disorders, and personality disorders. Studies that investigated strategies for alleviating caregiver burden were considered for inclusion. This also included studies that investigated the frameworks used to describe the burden of care among informal caregivers, such as caregiver strain, stress, and role fatigue.
Studies that reported on the burden of care without suggesting strategies for alleviating it were excluded.

**Context**

Studies from low- and middle-income countries, including countries in Africa, Asia, Latin America, and the Caribbean, were considered for inclusion. In addition, the inclusion criteria comprised studies from lower-middle-income countries using the World Bank classification of the economic status of a country at the time the study was conducted. The term developing country was included as it is similar to lower-middle-income country, and thus, countries in Africa, Asia, and Latin America (including the Caribbean) were eligible for inclusion.

**Types of Sources**

This review included published qualitative, quantitative, and mixed methods studies with all types of designs as well as unpublished studies, including dissertations and theses. Over the last decade, there has been a growing body of knowledge from studies outlining strategies for alleviating burden among informal caregivers in low- and middle-income countries. As a result, this review was only restricted to studies conducted between 2011 and 2021. Only papers written in English were included in this review.

**Search Strategy**

A 3-step search strategy process was used to identify relevant studies. The search strategy commenced with an initial limited search of PubMed, MEDLINE, CINAHL, and PsycINFO for published studies between 2011 and 2021. The keywords were as follows: "[informal caregiver/s OR caregiver/s] AND [caregiver burden OR caregiver stress] AND [support strategy/ies OR intervention/s] AND [severe mental disorder/s OR mental illness] AND [developing country/ies]."

The second search was refined with the assistance of the librarian at the University of the Witwatersrand, and additional terms were added: "[Carers or 'Informal Carers' or 'Male Caregiver' or 'Women Caregiver' or 'Family Caregivers'] AND [Caregiver* strain or 'Caregiver *Exhaustion' or 'Caregiver* burnout' or 'Carer* burden' or 'Carer* stress' or 'Caretaker* role fatigue' or 'Burden of Caregiver*' or 'Caretaker* burden' or 'Caregiver* load'] AND [Caregiver* Psychology]"

AND

"Strategy or 'Strategies' or 'Intervention' or 'Procedure' or 'Programme*' or 'Management' or 'Protocol* Guidelines' or 'Guide' or 'Policy' or 'Policies'"

AND

"Mental disorders' or 'Mental illness' or 'schizophrenia' or 'bipolar mood disorder' or 'Major depressive disorder' or 'Psychotic disorder' or 'Personality disorder' or 'Bipolar affective disorder'"

AND

(developing OR (less* N1 developed) OR "under developed") OR underdeveloped OR "under served" OR underserved OR deprived OR poor* OR "middle income" OR (low* N1 income)) N1 (country* OR nation* OR population* OR world) OR ((developing OR (less* N1 developed) OR "under developed") OR underdeveloped OR "under served" OR underserved OR deprived OR poor* OR "middle-income" OR (low* N1 income)) N1 (country* OR nation* OR population* OR world) OR ((developing OR (less* N1 developed) OR "under developed") OR underdeveloped)

**Source of Evidence Selection**

Following the search, all identified citations were exported to the Mendeley reference manager and thereafter to a web-based software, Covidence (Veritas Health Innovation), for primary screening and data extraction from the selected articles. Duplicates were removed, and 2 reviewers (OS and DC) independently conducted the screening for titles, abstracts, and full texts using the inclusion and exclusion criteria. To enhance the reliability of the review results, the first 5 abstracts were screened, and the 2 reviewers compared the screening results and clarified conflicts. Following this, the criteria were revised to ensure a detailed description of the various types of burden. The remaining abstracts were screened, and comparisons were made after 10 screenings to manage conflicts. The 2 reviewers (OS and DC) proceeded to screen texts and extract data from the selected articles. Any disagreements regarding the inclusion of the studies were resolved through discussion between the 2 reviewers.

**Data Extraction**

The Covidence software has templates for data extraction. These templates were modified using the data extraction tables proposed by Peters et al [30]. To ensure the extraction of
relevant data to answer the review question, these templates were piloted by the reviewers before use. In total, 2 separate data extraction templates (Multimedia Appendix 1) were used to extract data from studies that implemented an intervention and from studies that recommended interventions and strategies for alleviating caregiver burden. Data related to the characteristics of the studies were extracted, including study title; study aims; citation details; population of interest; concept of interest; context of the study, including the country and type of setting; type of evidence sources; study approach and designs; and participant characteristics, such as age, gender, and diagnosis of the care recipients. In addition, information on the characteristics of the interventions was extracted. This included the intervention content (ie, type of intervention, intervention developer and deliverer, and type of burden targeted by the intervention). The intervention description included the duration of the intervention, number of sessions, and location for the intervention. The templates for extracting the data are available in Multimedia Appendix 1. The reviewers extracted the data independently. Web-based meetings were held biweekly to discuss and resolve any discrepancies in the extracted data. Regular comparisons were easy and quick to conduct and improved the consistency of the extracted data. After completing the data extraction, the 2 reviewers scanned the references of the included articles to ensure that no articles were missed. Included and excluded studies were reported in a PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.
Data Analysis and Presentation
To increase consistency in the data, this review followed a 3-step analysis process as proposed by Levac et al [33]. This included analyzing the data, reporting the results, and applying meaning to the results [33]. The first step was reviewing the extracted data and identifying the type of data to be analyzed, which were then grouped according to the objectives of the review. This was followed by identifying the type of analysis appropriate for the specific type of data and then analyzing the extracted data. A descriptive quantitative analysis was conducted in a Microsoft Excel (Microsoft Corp) spreadsheet to describe the characteristics of the studies. This included the overall number of studies, types of study design, years of publication, characteristics of the study populations, countries where the studies were conducted, and types of interventions. In addition, descriptive content analysis was conducted deductively using the NVivo software (Lumivero) to code the characteristics of the extracted data into overall categories [34]. Data were categorized into types of strategies—implemented and recommended intervention strategies—and the outcomes of the strategies aimed at alleviating the burden of informal caregivers as reported by the authors. To enhance the clarity of the emerging findings, the data were summarized and presented in graphs and tables.

Results
Study Inclusion
A total of 18,323 studies were identified from the databases using the keywords. In total, 19 additional studies were identified from the reference lists of the included studies, thereby bringing the total number of potentially relevant studies to 18,342. Subsequently, of the 18,342 studies, 256 (1.4%) duplicates were removed, thereby leaving 18,086 (98.6%) studies deemed relevant to the review based on the keywords. A total of 98.83% (17,875/18,086) of these studies were then excluded; this exclusion of numerous studies occurred because of their incongruence with the predefined inclusion criteria, such as instances in which the study did not pertain to a low- or middle-income country context or in which the primary emphasis lay on the patient rather than the caregiver. Therefore, 1.17% (211/18,086) of the studies were screened by title and abstract based on set inclusion and exclusion criteria. This led to the exclusion of 58.8% (124/211) of these studies from the review. A total of 41.2% (87/211) of the studies were then retrieved as full texts based on eligibility by screening their titles and abstracts. Of these 87 studies, 43 (49%) were excluded for various reasons: 31 (72%) were not from low- or middle-income countries; 3 (7%) did not specify the intervention strategy; 2 (5%) focused on the wrong patient population; and another 7 (16%) were excluded for being a clinical guideline, focusing on patients rather than caregivers, having misaligned outcomes and a study design within the exclusion criteria (eg, review articles), being out of the date range of this review, the full text being in a language other than English, and a lack of an abstract and full text. Finally, 44 studies were included in this scoping review. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram was used to present the selection of the studies (Figure 1). It took approximately 8 months, from September 2021 to June 2022, to carry out the study from conception to completion of the project.

Characteristics of the Eligible Studies
Summary of Studies and Evidence Type
The included studies were published between 2011 and 2021. Most studies (31/44, 70%) were published between 2016 and 2021, and 30% (13/44) were published between 2011 and 2016. The types of studies included peer-reviewed journal articles (40/44, 91%), research reports or theses (2/44, 5%), and opinion pieces (2/44, 5%). Of the 44 included studies, 29 (66%) reported on a specific implemented intervention strategy, and 15 (34%) outlined recommended intervention strategies aimed at alleviating the burden of informal caregivers of people with severe and enduring mental health conditions in low- and middle-income countries. The study designs included randomized controlled trials (21/44, 48%), quantitative studies (15/44, 34%), qualitative studies (3/44, 7%), mixed methods studies (3/44, 7%), and gray literature (2/44, 5%; Table 1).
Table 1. Characteristics of the included studies (N=44).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Studies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary of studies</strong></td>
<td></td>
</tr>
<tr>
<td>Implemented strategy studies</td>
<td>29 (66)</td>
</tr>
<tr>
<td>Recommended strategy studies</td>
<td>15 (34)</td>
</tr>
<tr>
<td><strong>Type of evidence</strong></td>
<td></td>
</tr>
<tr>
<td>Journal articles</td>
<td>40 (91)</td>
</tr>
<tr>
<td>Research reports or theses</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Opinion pieces</td>
<td>2 (5)</td>
</tr>
<tr>
<td><strong>Year of publication</strong></td>
<td></td>
</tr>
<tr>
<td>2011-2015</td>
<td>13 (30)</td>
</tr>
<tr>
<td>2016-2021</td>
<td>31 (70)</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td></td>
</tr>
<tr>
<td>Randomized controlled trial</td>
<td>21 (48)</td>
</tr>
<tr>
<td>Quantitative design</td>
<td>15 (34)</td>
</tr>
<tr>
<td>Qualitative design</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Mixed methods</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Gray literature</td>
<td>2 (5)</td>
</tr>
</tbody>
</table>

**Context of the Studies**

A total of 44 studies identified from 16 low- and middle-income countries in Asia, Africa, Europe, and South and North America (Mexico) were included. The countries included China (12/44, 27%); Iran (10/44, 23%); India (5/44, 11%); Brazil (3/44, 7%); Greece (2/44, 5%); Ghana (2/44, 5%); and Thailand, Nepal, South Korea, Indonesia, Jordan, Mexico, Ethiopia, Turkey, Botswana, and mixed countries (including Brazil, Iran, Colombia, and Mexico 1/44, 2% of the studies each). As shown in Table 2, majority of the studies with implemented strategies (24/29, 83%) and those with recommended strategies (9/15, 60%) were from Asian countries. Overall, 7% (2/29) of the studies each from Europe, South America, and North America implemented the strategies for alleviating burden and 13% (2/15) of the studies from South and North America recommended the strategies. Only 3% (1/29) of the studies conducted in Africa implemented the strategies and 20% (3/15) of the studies recommended the strategies. A study (1/15, 7%) from mixed countries recommended the strategies for alleviating burden.

Table 2. Continents of the recommended and implemented strategies (N=44).

<table>
<thead>
<tr>
<th>Continent</th>
<th>Implemented strategies, n (%)</th>
<th>Recommended strategies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>24 (83)</td>
<td>9 (60)</td>
</tr>
<tr>
<td>Africa</td>
<td>1 (3)</td>
<td>3 (20)</td>
</tr>
<tr>
<td>Europe</td>
<td>2 (7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>South and North America</td>
<td>2 (7)</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Mixed countries</td>
<td>0 (0)</td>
<td>1 (7)</td>
</tr>
</tbody>
</table>

**Characteristics of the Study Populations**

The targeted populations in the included studies were mainly family caregivers (38/44, 86%), 7% (3/44) of the studies targeted informal carers, and some studies (3/44, 7%) did not specify their targeted population but rather identified them as caregivers. Most studies (33/44, 75%) included both male and female caregivers, and 25% (11/44) did not specify the sex of their sample. Similarly, most studies (24/44, 55%) included both male and female care recipients, only 2% (1/44) of the studies focused only on male care recipients, and 43% (19/44) did not specify the gender of the care recipients included in the studies. The diagnoses of the care recipients in the included studies were grouped into 6 categories: schizophrenia (17/44, 39%), mood disorders (4/44, 9%), multiple diagnoses including schizophrenia and mood disorders (5/44, 11%), mixed severe and enduring mental health conditions (5/44, 11%), other psychotic and chronic conditions (6/44, 14%), and unspecified mental disorders (7/44, 16%; Table 3).
Table 3. Characteristics of the study populations in the included studies (N=44).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Studies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population of interest</strong></td>
<td></td>
</tr>
<tr>
<td>Family caregiver or caregivers</td>
<td>38 (86)</td>
</tr>
<tr>
<td>Informal carer or carers</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Caregiver or caregivers</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Carer or carers</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Concept of interest</strong></td>
<td></td>
</tr>
<tr>
<td>Caregiver burden</td>
<td>18 (41)</td>
</tr>
<tr>
<td>Family burden</td>
<td>16 (36)</td>
</tr>
<tr>
<td>Caregiver stress</td>
<td>9 (20)</td>
</tr>
<tr>
<td>Caregiver strain</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Sex of caregivers</strong></td>
<td></td>
</tr>
<tr>
<td>Mixed-gender groups (male and female)</td>
<td>33 (75)</td>
</tr>
<tr>
<td>Not specified</td>
<td>11 (25)</td>
</tr>
<tr>
<td><strong>Gender of care recipients</strong></td>
<td></td>
</tr>
<tr>
<td>Mixed-gender groups (male and female)</td>
<td>24 (55)</td>
</tr>
<tr>
<td>Male</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Not specified</td>
<td>19 (43)</td>
</tr>
<tr>
<td><strong>Diagnosis of care recipients</strong></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>17 (39)</td>
</tr>
<tr>
<td>Mood disorders</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Schizophrenia and mood disorders</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Mixed diagnoses of severe mental disorders</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Unspecified mental disorders</td>
<td>7 (16)</td>
</tr>
<tr>
<td>Other (drug dependency, psychotic disorders, and chronic conditions)</td>
<td>6 (14)</td>
</tr>
</tbody>
</table>

**Concept of Interest**

The concepts of interest reported in the studies included caregiver burden (18/44, 41%), family burden (16/44, 36%), caregiver stress (9/44, 20%), and caregiver strain (1/44, 2%; Table 3).

**Review Findings**

This section outlines the findings of the scoping review and includes the types of intervention strategies, characteristics of the intervention strategies, and author-reported outcomes from the included studies.

**Types of Intervention Strategies**

The various strategies for alleviating caregiver burden were grouped into 2 categories: implemented and recommended intervention strategies. The implemented intervention strategies emerged from studies that investigated the effects of a specific intervention to alleviate the burden of informal caregivers. The recommended interventions were identified from studies that investigated caregiver burden and outlined recommended intervention strategies to be used to alleviate the burden of informal caregivers of persons with severe and enduring mental health conditions. Most studies (29/44, 66%) reported on an implemented intervention strategy, and 34% (15/44) of the studies recommended strategies to alleviate caregiver burden. Most of the studies and articles (33/44, 75%) originated in Asia, with 55% (24/44) implemented and 20% (9/44) recommended intervention strategies reported on, followed by Africa with 2% (1/44) implemented and 7% (3/44) recommended intervention strategies, South and North America with 5% (2/44) implemented and 5% (2/44) recommended intervention strategies, and Europe with 5% (2/44) implemented intervention strategies (Table 2). A study from mixed countries only reported recommended intervention strategies. The identified strategies were grouped into the following categories: community-based interventions, psychoeducation interventions, support groups, cognitive behavioral therapy (CBT), spirituality-based interventions, smartphone-based interventions, mindfulness and empowerment, collaborative interventions, standard care, financial and social support, counseling, occupation-based interventions, policy and legislature, and access to mental health care (Table 4 and Textbox 2). These categories are discussed in the following subsections.
<table>
<thead>
<tr>
<th>Type of strategy</th>
<th>Evaluated interventions</th>
<th>Author-reported outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community-based interventions</td>
<td>CoMHIP&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Positive outcomes (positive effects on family caregivers)</td>
</tr>
<tr>
<td></td>
<td>RESHAPE&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Psychoeducation interventions</td>
<td>Psychoeducation program</td>
<td>Improved coping skills</td>
</tr>
<tr>
<td></td>
<td>Group psychoeducation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FLEP&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Peer-assisted education</td>
<td></td>
</tr>
<tr>
<td>Support groups</td>
<td>FPGP&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Knowledge and skill acquisition</td>
</tr>
<tr>
<td></td>
<td>Family-led mutual support program</td>
<td></td>
</tr>
<tr>
<td>Cognitive behavioral therapy</td>
<td>Emotional regulation training</td>
<td>Reduced burden of care</td>
</tr>
<tr>
<td>Spirituality-based intervention</td>
<td>Spirituality-based program</td>
<td>Reduced anxiety and stress</td>
</tr>
<tr>
<td>Guided self-help interventions</td>
<td>Manual-guided PBS&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Strengthened positive caregiving experience</td>
</tr>
<tr>
<td></td>
<td>The Good Mood Guide</td>
<td>Reduced negative caregiving experience</td>
</tr>
<tr>
<td>Smartphone-based interventions</td>
<td>MHapps&lt;sup&gt;h&lt;/sup&gt;</td>
<td>Increased access to support</td>
</tr>
<tr>
<td>Mindfulness and empowerment</td>
<td>IEP&lt;sup&gt;i&lt;/sup&gt;</td>
<td>Caregiver empowerment</td>
</tr>
<tr>
<td>interventions</td>
<td>MBSR&lt;sup&gt;j&lt;/sup&gt; program</td>
<td>Improved QoL</td>
</tr>
<tr>
<td>Collaborative interventions</td>
<td>Participatory care model</td>
<td>Reduced burden of care</td>
</tr>
<tr>
<td>Standard care</td>
<td>Multimodal intervention (general medicine, psychiatry, psychology, family therapy, neuropsychological rehabilitation, and occupational therapy)</td>
<td>Reduced burden of care</td>
</tr>
</tbody>
</table>

<sup>a</sup>CoMHIP: community mental health early intervention project.
<sup>b</sup>RESHAPE: Reducing Stigma Among Healthcare Providers.
<sup>c</sup>FLEP: Family Link Education Programme.
<sup>d</sup>QoL: quality of life.
<sup>e</sup>BMD: bipolar mood disorder.
<sup>f</sup>FPGP: Family-led Peer Support Group Program.
<sup>g</sup>PBS: problem-solving–based self-learning program.
<sup>h</sup>MHapps: mental health apps.
<sup>i</sup>IEP: integrated empowerment program.
<sup>j</sup>MBSR: mindfulness-based stress reduction.
**Textbox 2. Description of the identified recommended strategies for alleviating informal caregiver burden.**

<table>
<thead>
<tr>
<th>Community-based interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Strengthening existing primary health care system</td>
</tr>
<tr>
<td>• Establishment of day nursing or care</td>
</tr>
<tr>
<td>• Respite care</td>
</tr>
<tr>
<td>• Vocational training for patients</td>
</tr>
<tr>
<td>• Single-family caregiver organization</td>
</tr>
<tr>
<td>• Supported employment for patients</td>
</tr>
<tr>
<td>• Home visits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Psychoeducation interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Group psychoeducation</td>
</tr>
<tr>
<td>• Family education</td>
</tr>
<tr>
<td>• Continuing education programs for effective practice</td>
</tr>
<tr>
<td>• Educational lectures</td>
</tr>
<tr>
<td>• Psychoeducation program</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Support groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Nongovernmental mental health–related support groups</td>
</tr>
<tr>
<td>• Family caregiver assistance programs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cognitive behavioral therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Family therapy groups</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spirituality-based intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Turning to traditional healers and spiritual leaders</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smartphone-based interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Mental health apps</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mindfulness and empowerment interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Stress management training</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collaborative interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Caregiver involvement in all program elements</td>
</tr>
<tr>
<td>• Pluralistic and ecological approach to service delivery</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard care</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Telepsychiatry</td>
</tr>
<tr>
<td>• Availability of emergency teams</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Financial support</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Disability grant</td>
</tr>
<tr>
<td>• Fee-free mental health services</td>
</tr>
<tr>
<td>• Medical insurance and free medication</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social support</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Medical care social security</td>
</tr>
<tr>
<td>• Social resources</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Counseling</th>
</tr>
</thead>
</table>

https://www.i-jmr.org/2024/1/e48587

Interact J Med Res 2024 | vol. 13 | e48587 | p.49

(page number not for citation purposes)
Basic counseling
Supportive psychotherapy

**Occupation-based interventions**
Physical and leisure activities

**Policy and legislature**
National health insurance schemes
Advocacy for a strong mental health policy
Integration of caregiver actions and interventions into national mental health care plans
Fee-free mental health services

**Access to mental health care**
Caregiver care inclusion in daily treatment facilities
Access to therapeutic tools
Availability of mental health units or departments
Periodic health checks for caregivers (every 6 months on average)
Improvement in institutional mental health care

**Psychoeducational Interventions**
The first strategy was psychoeducational interventions, which was identified as the most common strategy implemented and recommended to alleviate caregiver burden. A total of 27% (12/44) of the studies implemented 4 different psychoeducational interventions, including 58% (7/12) of these studies using a psychoeducation program [35-40], 17% (2/12) using group psychoeducation [41,42], 17% (2/12) using the Family Link Education Programme [43,44], and 8% (1/12) of the studies implementing peer-assisted education [45] (Table 4). Similarly to the implemented intervention strategies, the recommended psychoeducational interventions identified included psychoeducational programs, group psychoeducation, family psychoeducation, continuing education programs for effective practice, and educational lectures [7,46] (Textbox 2).

**Support Group Interventions**
The second strategy was support groups. In total, 7% (3/44) of the studies implemented support group strategies, including the Family-led Peer Support Group Program, family-led mutual support programs, and nongovernmental mental health–related support groups [47-49] (Table 4). The recommended support group strategies identified in this review included nongovernmental mental health–related support groups and family caregiver support programs [50-54] (Textbox 2).

**Community-Based Interventions**
Community-based interventions were identified as a third strategy to alleviate caregiver burden, and 5% (2/44) of the included studies reported on implementing a community mental health early intervention project and Reducing Stigma Among Healthcare Providers [55,56] as community-based interventions (Table 4). The recommended community-based interventions included strengthening existing primary health care systems, establishing day nursing or care for the care recipients, respite care, vocational rehabilitation training for patients, single-family caregiver organizations, supported employment for patients, and home visits [46,52,53,57] (Textbox 2).

**Guided Self-Help Interventions**
A total of 5% (2/44) of the studies implemented guided self-help interventions, namely, a manual-guided problem-solving–based self-learning program [58] and the Good Mood Guide [59] (Table 4). No recommended guided self-help interventions were identified in this review.

**Mindfulness and Empowerment Interventions**
Mindfulness and empowerment interventions were implemented in 5% (2/44) of the studies, and the programs implemented were a mindfulness-based stress reduction program and an integrated empowerment program [60,61] (Table 4). Stress management training was identified as a recommended mindfulness and empowerment intervention strategy in this review [7] (Textbox 2). The recommended cognitive behavioral strategy identified in this review was family therapy groups [64] (Textbox 2).

**CBT Interventions**
Emotional regulation training was the only implemented CBT strategy identified in 5% (2/44) of the included studies [62,63] (Table 4). The recommended cognitive behavioral strategy identified in this review was family therapy groups [64] (Textbox 2).

**Spirituality-Based Interventions**
In total, 2% (1/44) of the studies implemented a spirituality-based program for informal caregivers in which the comparison intervention was 2 standard group training sessions related to general mental disorders [65] (Table 4). Turning to traditional healers and spiritual leaders was identified as a recommended spirituality-based intervention for alleviating informal caregiver burden [52] (Textbox 2).
Participatory Care Model
In total, 2% (1/44) of the studies implemented a participatory care model as a collaborative intervention strategy [66] (Table 4). The recommended collaborative interventions were caregiver involvement in all program elements and a pluralistic and ecological approach to service delivery [56,67] (Textbox 2).

Standard Care
The implemented standard care interventions were multimodal interventions that encompassed general medicine, psychiatry, psychology, family therapy, neuropsychological rehabilitation, and occupational therapy with no comparison intervention [68] (Table 4). The recommended standard care interventions were telepsychiatry and ensuring the availability of emergency teams [53] (Textbox 2).

Smartphone-Based Interventions
Only 2% (1/44) of the studies implemented smartphone-based interventions using mental health apps to alleviate the burden of informal caregivers with no comparison intervention [67] (Table 4). Although the study indicated the use of the mental health apps with care recipients and their caregivers, the specific intervention offered was not described. No recommended smartphone-based interventions were identified in this review.

Additional Strategies Recommended for Alleviating Caregiver Burden
The additional recommended intervention strategies were financial support, including the provision of a disability grant, fee-free mental health services, medical insurance, and free medication for care recipients [7,57,68]. The recommended social support comprised medical care social security and social resources [52,54,56,57,69]. Recommended counseling interventions comprised basic counseling and supportive psychotherapy [64,69]. Occupation-based interventions included participation in physical and leisure activities [64]. Policy and legislature recommended interventions comprised the implementation of national health insurance schemes, advocacy for a strong mental health policy, integration of caregiver actions and interventions into national mental health care plans, and fee-free mental health services [53,62,68]. Finally, other recommended interventions were access to mental health care, which considers caregiver care inclusion in daily treatment facilities; access to therapeutic tools; availability of mental health units or departments; periodic health checks for caregivers (every 6 months on average); and improved institutional mental health care [35,52,54,67] (Textbox 2).

Characteristics of the Implemented Intervention Strategies
Most of the implemented intervention strategies (18/29, 62%) were targeted at alleviating both objective and subjective burden, and 38% (11/29) were exclusively aimed at alleviating subjective burden. The intervention developers were reported as being researchers based on evidence (13/29, 45%) and researchers with expert input (5/29, 17%). In total, 28% (8/29) were existing interventions implemented without any adaptations, and 10% (3/29) of the interventions did not specify the intervention developer. The intervention deliverers identified in the studies included trained peer facilitators (7/29, 24%), researchers (7/29, 24%), psychiatrists and nurses (5/29, 17%), multiple health care professionals (4/29, 14%), and psychologists (3/29, 10%), and 10% (3/29) of the interventions did not specify who delivered them.

The number of sessions offered in the implemented intervention strategies was between 6 and 8 sessions (13/29, 45%), ≤5 sessions (8/29, 28%), and ≥11 sessions (5/29, 17%), and 10% (3/44) of the interventions did not specify the number of sessions. The duration of the implemented interventions in the studies was 6 to 15 weeks (10/29, 34%), ≤5 weeks (7/29, 24%), and ≥16 weeks (4/29, 14%), and 28% (8/29) of the interventions did not specify the duration. Most interventions (9/29, 31%) were delivered once per week, some were offered twice a week (8/29, 28%), and some of the interventions (12/29, 41%) did not specify the frequency of their sessions. The length of the sessions in the implemented interventions ranged between 2 and 2.5 hours (8/29, 28%), 1 to 1.5 hours (6/29, 21%), and ≤1 hour (1/29, 3%), and 48% (14/29) of the interventions did not specify the length of their sessions (Table 5).
Table 5. Characteristics of the implemented strategies for alleviating informal caregiver burden (N=29).

<table>
<thead>
<tr>
<th>Study</th>
<th>Year of publication</th>
<th>Type of intervention</th>
<th>Type of burden targeted by the intervention</th>
<th>Intervention developer</th>
<th>Intervention deliverer</th>
<th>Frequency of sessions per week</th>
<th>Duration of intervention</th>
<th>Number of sessions</th>
<th>Length of the sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ng et al [56]</td>
<td>2018</td>
<td>CoMHIP(^a)</td>
<td>Subjective and objective</td>
<td>Not specified</td>
<td>The Integrated Community Centre for Mental Wellness</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Rai et al [55]</td>
<td>2018</td>
<td>RESHAPE(^b)</td>
<td>Subjective and objective burden</td>
<td>Developed as part of WHO(^c) mhGAP(^d) and PRIME(^e)</td>
<td>TPO(^f) Nepal, a Nepali non-governmental organization</td>
<td>5 sessions</td>
<td>4 days</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Rajai et al [45]</td>
<td>2021</td>
<td>Peer-assisted education</td>
<td>Subjective and objective burden</td>
<td>Researchers and approved by 3 faculty members of the Army College of Medical Sciences</td>
<td>Trained peer facilitators</td>
<td>6 sessions</td>
<td>3 weeks</td>
<td>Twice weekly</td>
<td>1 hour</td>
</tr>
<tr>
<td>Zhou et al [44]</td>
<td>2020</td>
<td>FLEP(^g)</td>
<td>Subjective and objective burden</td>
<td>FLEP: peer-led psychoeducation program developed from the stress and coping model by Pearlin et al [70]</td>
<td>Group facilitators who were peer specialists of experienced caregivers</td>
<td>8 sessions</td>
<td>8 weeks</td>
<td>Once weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Dewi et al [37]</td>
<td>2019</td>
<td>Family psychoeducation and care decision without pasung</td>
<td>Subjective and objective burden</td>
<td>Researchers</td>
<td>Researchers</td>
<td>3 sessions</td>
<td>3 weeks</td>
<td>Not specified</td>
<td>35-40 minutes</td>
</tr>
<tr>
<td>Tabeleão et al [36]</td>
<td>2018</td>
<td>Psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Not specified</td>
<td>10 psychologists</td>
<td>6 sessions</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Ntsayagae [40]</td>
<td>2017</td>
<td>Psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Researcher based on existing literature</td>
<td>Researcher</td>
<td>2 sessions</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>De Souza et al [39]</td>
<td>2016</td>
<td>Psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Psychoeducational intervention for patients with BD(^h) by Colom et al [71] translated and adapted by Dell-Aglio et al [72]</td>
<td>Psychiatrist responsible for specific case</td>
<td>6 sessions</td>
<td>6 to 8 weeks</td>
<td>Twice weekly</td>
<td>Not specified</td>
</tr>
<tr>
<td>Kolostoumpis et al [38]</td>
<td>2015</td>
<td>Psychoeducation</td>
<td>Subjective burden</td>
<td>Adapted from the treatment protocol developed by Reinares et al [73] in the Barcelona Bipolar Disorders Program in Spain</td>
<td>Psychiatrist and psychologist</td>
<td>7 sessions</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Study</td>
<td>Year of publication</td>
<td>Type of intervention</td>
<td>Type of burden targeted by the intervention</td>
<td>Intervention developer</td>
<td>Intervention deliverer</td>
<td>Number of sessions</td>
<td>Duration of intervention</td>
<td>Frequency of sessions per week</td>
<td>Length of the sessions</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>---------------------------------------------</td>
<td>------------------------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>-------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Fallahi Khoshknab et al [41]</td>
<td>2014</td>
<td>Group psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Researchers based on educational program of psychiatric nursing and psychiatric textbooks</td>
<td>Organized based on educational program of psychiatric nursing and psychiatric textbooks (Campbell [74]) and converted to understandable text for patients and families</td>
<td>4 sessions</td>
<td>4 weeks</td>
<td>Once weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Navidian et al [42]</td>
<td>2012</td>
<td>Psychoeducation</td>
<td>Subjective burden</td>
<td>Researchers based on the families’ needs and the existing literature</td>
<td>Psychiatrist and mental health nurse</td>
<td>4 sessions</td>
<td>4 weeks</td>
<td>Once weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Sharif et al [75]</td>
<td>2012</td>
<td>Psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Psychiatrist and psychiatric nurse based on the literature and needs of the families</td>
<td>Psychiatrist, nurse, and guest speakers</td>
<td>10 sessions</td>
<td>5 weeks</td>
<td>Not specified</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Chiu et al [43]</td>
<td>2011</td>
<td>FLEP</td>
<td>Subjective burden</td>
<td>Task force consisting of a mental health social worker, a recovered patient with editorial experience, and a caregiver based on available related local educational materials and the NAMI Family-to-Family program</td>
<td>Trainers who were themselves family members of people with SMI. They received training and a trainer manual.</td>
<td>8 sessions</td>
<td>Not specified</td>
<td>Once weekly</td>
<td>Not specified</td>
</tr>
<tr>
<td>Chien et al [48]</td>
<td>2018</td>
<td>Family-led support program</td>
<td>Subjective and objective burden</td>
<td>Family-led mutual support group—contents were based on similar program protocols and the researcher-developed family mutual support groups for psychotic disorders</td>
<td>Family-led mutual support group—co-led by 2 peer family caregivers along with a researcher and rehabilitation nurse</td>
<td>16 sessions</td>
<td>36 weeks</td>
<td>Twice weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Chien et al [48]</td>
<td>2018</td>
<td>Family-led mutual support program</td>
<td>Subjective burden</td>
<td>6 experts on psychiatric rehabilitation (including psychiatrists, clinical psychologists, and nurse specialists)</td>
<td>Peer leader who received training from the researchers and worked closely with a group leader who was a trained psychiatric nurse</td>
<td>14 sessions</td>
<td>39 weeks</td>
<td>Twice weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Study</td>
<td>Year of publication</td>
<td>Type of intervention</td>
<td>Type of burden targeted by the intervention</td>
<td>Intervention developer</td>
<td>Intervention duration</td>
<td>Number of sessions</td>
<td>Duration of intervention</td>
<td>Frequency of sessions per week</td>
<td>Length of the sessions</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>---------------------------------------------</td>
<td>------------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>-------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Chien and Chan [47]</td>
<td>2013</td>
<td>FPGP&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Subjective and objective burden</td>
<td>Content and format were based on previous programs conducted by Li and Arthur [76] in mainland China. Appropriateness of the content was rated by 7 experts, including psychiatrists, psychologists, and nursing specialists.</td>
<td>Peer support—peer leader supported by principal researcher</td>
<td>14 group sessions</td>
<td>39 weeks</td>
<td>Twice weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Behrouian et al [62]</td>
<td>2021</td>
<td>Emotional regulation</td>
<td>Subjective burden</td>
<td>On the basis of the Dialectical Behavior Therapy Skills Workbook and CBT&lt;sup&gt;m&lt;/sup&gt; principles</td>
<td>Clinical psychologist</td>
<td>8 sessions</td>
<td>8 weeks</td>
<td>Once weekly</td>
<td>Not specified</td>
</tr>
<tr>
<td>Behrouian et al [63]</td>
<td>2020</td>
<td>Emotion regulation training</td>
<td>Subjective and objective burden</td>
<td>Sessions were based on previous studies (Gratz and Gunderson [77]). Trainings were based on the Dialectical Behavior Therapy Skills Workbook (McKay et al [78]).</td>
<td>Clinical psychologist</td>
<td>8 sessions</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Faghih and Pahalvanzadeh [79]</td>
<td>2019</td>
<td>CBT</td>
<td>Subjective burden</td>
<td>Researchers based on existing literature</td>
<td>Researcher</td>
<td>16 sessions</td>
<td>8 weeks</td>
<td>Twice weekly</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Khosravi et al [65]</td>
<td>2021</td>
<td>Spirituality-based program</td>
<td>Subjective burden</td>
<td>Developed by McCann et al [80] in Australia, and its Chinese translated version was validated and refined by the research team.</td>
<td>Psychiatric nurse</td>
<td>5 modules</td>
<td>21 weeks</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Chien et al [58]</td>
<td>2020</td>
<td>Manual-guided PBSP&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Subjective and objective burden</td>
<td>The Good Mood Guide developed by Lifeline South Coast NSW&lt;sup&gt;o&lt;/sup&gt;, Australia</td>
<td>Researcher via telephone</td>
<td>8 modules</td>
<td>8 weeks</td>
<td>Once weekly</td>
<td>Not specified</td>
</tr>
<tr>
<td>McCann et al [59]</td>
<td>2015</td>
<td>Guided self-help manual</td>
<td>Subjective and objective burden</td>
<td>The Good Mood Guide developed by Lifeline South Coast NSW&lt;sup&gt;o&lt;/sup&gt;, Australia</td>
<td>Researcher</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Deb et al [67]</td>
<td>2018</td>
<td>Smartphone-based interventions (MHapps&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>Subjective and objective burden</td>
<td>Researchers based on existing literature</td>
<td>Psychiatrist</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Study</td>
<td>Year of publication</td>
<td>Type of intervention</td>
<td>Type of burden targeted by the intervention</td>
<td>Intervention developer</td>
<td>Intervention duration</td>
<td>Number of sessions</td>
<td>Duration of intervention</td>
<td>Frequency of sessions per week</td>
<td>Length of the sessions</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>---------------------------------------------</td>
<td>-------------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Hyun et al [61]</td>
<td>2018</td>
<td>IEP&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Subjective burden</td>
<td>Developed by Hyun et al [61] for community-living PMIs&lt;sup&gt;b&lt;/sup&gt; based on the empowerment theories of Kanter [81] and McLean [82]</td>
<td>Mental health professionals who received training with a written structured intervention protocol from the research team</td>
<td>4 sessions</td>
<td>4 weeks</td>
<td>Once weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Hou et al [60]</td>
<td>2014</td>
<td>MBSR&lt;sup&gt;c&lt;/sup&gt; pro-</td>
<td>Subjective burden</td>
<td>Not stated but assumed that the researchers modeled the program on the original MBSR by Kabat-Zinn [83]</td>
<td>3 trained instructors with &gt;3 years of experience in MBSR</td>
<td>8 sessions</td>
<td>8 weeks</td>
<td>Once weekly</td>
<td>2 hours, with a CD of 30-45 minutes for home practice</td>
</tr>
<tr>
<td>Zoladl et al [66]</td>
<td>2020</td>
<td>Participatory care model</td>
<td>Subjective burden</td>
<td>Researchers</td>
<td>Researchers and staff at the research site</td>
<td>8 sessions</td>
<td>12 weeks</td>
<td>Once weekly</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Ramirez et al [68]</td>
<td>2017</td>
<td>Standard care—included care from general medicine, psychiatry, psychology, family therapy, neuropsychological rehabilitation, and occupational therapy</td>
<td>Subjective and objective burden</td>
<td>Not stated</td>
<td>Medical officer, psychiatrist, psychologist, and occupational therapist</td>
<td>12-18 sessions</td>
<td>10 weeks</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
</tbody>
</table>

<sup>a</sup>CoMHiP: community mental health early intervention project.
<sup>b</sup>RESHAPE: Reducing Stigma Among Health Care Providers.
<sup>c</sup>WHO: World Health Organization.
<sup>d</sup>mhGAP: Mental Health Gap Action Programme.
<sup>e</sup>PRIME: Programme for Improving Mental Health Care.
<sup>f</sup>TPO: Transcultural Psychosocial Organization.
<sup>g</sup>FLEP: Family Link Education Programme.
<sup>h</sup>BD: bipolar disorder.
<sup>i</sup>NAMI: National Alliance on Mental Illness.
<sup>j</sup>SMI: severe mental illness.
<sup>k</sup>NGOMH: nongovernmental mental health.
<sup>l</sup>FPGP: Family-led Peer Support Group Program.
<sup>m</sup>CBT: cognitive behavioral therapy.
<sup>n</sup>PBSP: problem-solving–based self-learning program.
<sup>o</sup>NSW: New South Wales.
<sup>p</sup>MHapps: mental health apps.
<sup>q</sup>IEP: integrated empowerment program.
<sup>r</sup>PMI: person with mental illness.
<sup>s</sup>MBSR: mindfulness-based stress reduction.
**Perceived Effectiveness of the Implemented Intervention Strategies**

This review outlines the effectiveness of the implemented intervention strategies as reported by the authors who conducted the studies (Table 4). It should be noted that this scoping review included both published and gray literature, and therefore, no critical appraisal of the studies or meta-analyses were conducted. Therefore, care should be taken as these interpretations may be clouded by author bias [84].

The authors expressed the effectiveness of the implemented intervention strategies by highlighting whether the strategy resulted in a positive or negative outcome for the informal caregivers or their care recipients. Overall, the implemented strategies were reported to be effective in reducing the burden of care and improving the quality of life of informal caregivers. Psychoeducation intervention strategy outcomes included improved coping skills, improved quality of life, reduced anxiety and stress, reduced burden of care, reduced psychological strain, improved knowledge and skills in caregiving, enhanced family functioning, enhanced recovery, relapse prevention, and reduced prolonged admissions. Support group outcomes included alleviation of guilt, enhanced help-seeking behavior, improved quality of life, improved knowledge and skills in caregiving, improved psychosocial health, improved family and patient functioning, and decreased demand for mental health services. CBT outcomes included reduction in burden of care, anxiety, and stress and increased resilience of informal caregivers. Guided self-help outcomes included strengthened positive caregiving experiences, reduced negative caregiving experiences, and increased access to support. The reported outcomes for mindfulness and empowerment interventions included empowerment of caregivers, increased self-efficacy, improved quality of life, and decreased depressive symptoms. Collaborative intervention outcomes included reduced burden of care and increased resilience. The standard care outcomes reported included positive outcomes in caregiving, reduction in the burden of care, and increase in social support for caregivers.

Negative outcomes were only reported for the psychoeducation and mindfulness and empowerment intervention strategies. The reported negative outcomes of psychoeducation interventions were no improvement in the burden of care, quality of life, or self-esteem. The reported negative outcomes of mindfulness and empowerment interventions included short-lived improvement in anxiety, and the authors highlighted no improvement after 3 months of follow-up with the participants and no effect on perceived stress, quality of life, and self-compassion, indicating that the effects of the interventions were not sustainable.

**Discussion**

**Principal Findings**

This scoping review set out to map the literature on strategies for alleviating the burden of informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. The review identified types of strategies, strategy characteristics, and outcomes of the strategies as reported by the authors. The types of strategies identified were categorized as implemented and recommended intervention strategies and included community-based interventions, psychoeducational interventions, support groups, CBT, spirituality-based interventions, guided self-help, smartphone-based interventions, mindfulness and empowerment, collaborative interventions, standard care, financial and social support, counseling, occupation-based interventions, policy and legislature, and access to mental health.

Most of the implemented and recommended intervention studies (33/44, 75%) were conducted in Asian countries and targeted both subjective and objective burdens, and some studies (18/29, 62%) focused exclusively on subjective burden. This shows an increasing research interest in caregiver interventions in mental health research in Asian countries over the last decade. In contrast, the small number of studies from other low- and middle-income countries possibly confirms the limited research on support strategies for alleviating caregiver burden in mental health. Despite consensus on the high levels of burden of informal caregivers of persons with severe and enduring mental health conditions in these countries [56,85-87] and the importance of supporting caregivers, few studies (44/18,086, 0.24%) implemented and recommended intervention strategies for alleviating caregiver burden. This evidence gap raises significant concerns considering the scarcity of mental health care professionals and limited access to quality mental health services in many low- and middle-income countries. As a result, informal caregivers assume a vital role in the support and care of persons with severe and enduring mental health conditions in these regions [88,89]. Having limited strategies for alleviating the burden of informal caregivers may have dire consequences for the management of persons with severe and enduring mental health conditions in these countries.

The implemented and recommended intervention strategies were mainly focused on both male and female family caregivers. These findings are consistent with the literature as families have long been acknowledged as key stakeholders in the care and management of mental disorders [90-93]. Therefore, it is important to ensure that the strategies for alleviating caregiver burden are targeted specifically to this population, particularly in low- and middle-income countries, where occupying this role is often obligatory [94]. The fact that most implemented intervention strategies were aimed at both male and female caregivers is important to note. Although caregiving is identified as a female-oriented role, there is evidence that men also occupy the primary caregiver role and are also likely to experience high levels of caregiver burden [20,94]. This emphasizes the need to ensure that, where strategies for alleviating caregiver burden are implemented, they focus on all caregivers irrespective of their gender. The care recipients in these studies were mostly diagnosed with schizophrenia and mood disorders. This is consistent with the literature highlighting that depressive disorders, schizophrenia, and bipolar disorders are among the top 10 leading causes of disability in low- and middle-income countries [95].

The findings of this scoping review emphasize the need for evidence-based intervention strategies aimed at alleviating the
burden of informal caregivers in low- and middle-income countries. Most implemented intervention strategies were informed by evidence, in which researchers consulted existing literature and sought expert input to develop their interventions. Trained peer facilitators delivered most of the implemented interventions and were informal caregivers themselves. This is important to note as it aligns with task shifting, which is focused on transferring skills and responsibilities to local people with the aim of increasing access to mental health services in low- and middle-income countries where there is a shortage of human resources [96,97]. In addition, the findings of this review reveal that researchers, nurses, and psychiatrists offered some of the interventions, which may not be sustainable in low- and middle-income countries given the shortage of mental health professionals [95]. Most studies (13/29, 45%) offered 6 to 8 sessions over 6 to 8 weeks, and the sessions were facilitated once or twice a week for 1 to 2 and a half hours. Given the number and frequency of sessions, it may be useful to use peer facilitators to offer these interventions as most are already in the community, which will ensure sustainability in resource-constrained contexts in low- and middle-income countries.

Overall, the implemented intervention strategies were reported to have a positive effect on alleviating the burden of informal caregivers of people with severe and enduring mental health conditions in low- and middle-income countries. In some studies, these effects were reported for both informal caregivers and their care recipients. The authors reported the effectiveness of the strategies in their studies in terms of the help, benefit, and effect that a specific implemented intervention strategy had on the informal caregivers. Some compared their interventions mainly with standard psychiatric care and psychoeducation, which focused on providing caregivers with information on specific or general mental disorders. Psychoeducational interventions were frequently implemented intervention strategies and were identified as helpful, beneficial, and effective in reducing caregiver burden [35,38,42,75]. This type of intervention was reported to improve caregiver knowledge and skills to enable them to cope with the demands of caregiving. In addition, psychoeducational interventions were linked with positive patient outcomes as they were reported to enhance patient recovery, which subsequently reduced relapses and prolonged admissions. These interventions were commended for being simple, feasible, and cost-effective, making them the most preferred form of intervention strategy to address the burden of informal caregivers.

Similarly, implemented intervention strategies such as support groups, community-based interventions, guided self-help, mindfulness and empowerment, and CBT were reported to have a positive effect on the burden of informal caregivers in low- and middle-income countries. It is important to note that the family-led mutual support program, a peer-facilitated intervention, was reported to have long-term desirable benefits on the psychosocial health of both caregivers and their care recipients compared with the psychoeducational program and standard care offered by psychiatrists, clinical psychologists, and nursing specialists. Similarly, the manual-guided problem-solving–based self-learning program, a guided self-help intervention, was reported to have a superior treatment effect on caregiver burden, care recipients’ symptom severity, and the duration of rehospitalizations at the 6-month follow-up compared with a well-accepted family psychoeducation group program. This implies that support groups and guided self-help, although not frequently implemented, should be considered beneficial strategies for alleviating caregiver burden. In addition, as self-directed and peer-evaluated interventions, these intervention strategies may be considered cost-effective in extending the services to informal caregivers, especially in low- and middle-income countries where a shortage of human resources affects the delivery and quality of mental health services.

Although not frequently implemented, intervention strategies such as spirituality-based, collaborative, and standard care interventions were also reported to reduce caregiver burden. Spirituality-based interventions were identified as an inexpensive and readily available resource for alleviating the burden of informal caregivers. Similarly, the participatory care model was reported to be an efficient and low-cost method for reducing caregiver burden and increasing caregiver resilience. Therefore, it is necessary to further explore the use of these interventions with informal caregivers of people with severe and enduring mental health conditions in low- and middle-income countries.

Despite their potential, the use of virtual interventions or telehealth strategies for informal caregivers remains underrepresented in the literature. This may be attributed to the relatively slow adoption of telehealth among informal caregivers. Notably, only 1 smartphone app was identified in this scoping review. However, telehealth holds promise as a viable approach for disseminating psychoeducational information, providing real-time support, and facilitating participation in virtual caregiver support groups, particularly in remote areas. The increasing proliferation of smartphones and internet access in Africa further emphasizes opportunities for the development of telehealth programs tailored to caregivers of individuals with severe and enduring mental health conditions. Although most of the implemented intervention strategies reported positive effects on caregiver burden, there were negative outcomes reported for some interventions. Hou et al [60] reported that the mindfulness-based stress reduction program led to a short-lived improvement in the stress and anxiety of the informal caregivers but that no improvement was noted after the 3-month follow-up. In addition, this intervention did not demonstrate a sustained effect on health-related quality of life, perceived stress, or self-compassion. These outcomes were attributed to the loss to follow-up and the fact that the measurement instruments could have been less sensitive to changes in quality of life and stress. Similarly, the 6-session individual psychoeducation intervention for informal caregivers of persons with bipolar mood disorder showed no effect on burden, quality of life, or self-esteem. This was attributed to the reduction in the number of sessions from 21 in the original instruction [71] to 6 in this study [35]. Another reason was that the sessions were conducted individually, whereas the literature highlights that multifamily intervention groups are effective [71,73,98].
Implications for Practice and Policy

The findings of this review provide evidence for the reported evaluated and recommended interventions having benefits in reducing the burden of care among informal caregivers in low- and middle-income countries. Previous studies conducted in low- and middle-income countries have since emphasized the urgent need to strengthen informal caregiver support regarding mental health. The need for information and skills in handling mental health care users, as well as the need for emotional and tangible support, has been highlighted in previous studies [42,99]. The findings of this review revealed that psychoeducation and support groups were highly used strategies for improving knowledge and skills as well as building support for informal caregivers in low- and middle-income countries. Furthermore, these strategies were reported to be beneficial and cost-effective, thereby making them a viable option for implementation in low- and middle-income countries, where limited access to mental health resources prevails [100]. The need to ensure that these strategies are offered on a continuous basis was highlighted in many studies, indicating the need for 6 to 10 sessions over a period of 6 to 10 weeks [7]. To ensure sustainability, training of peer facilitators to deliver these interventions may be realistic as this aligns with a task-shifting approach, which has long been advocated for as a cost-effective strategy for increasing access to mental health services in low- and middle-income countries [101]. It is interesting to note that peer support groups and guided self-help were reported to have long-term benefits on caregiver burden compared with psychoeducation; this highlights that using self-directed interventions is considered a practical option for alleviating informal caregiver burden. In low- and middle-income countries where a shortage of human resources prevails, it is important to consider such interventions as they empower informal caregivers to build support and take ownership of their health and well-being [35]. Spiritual-based interventions, although not frequently implemented, are important to note in the context of low- and middle-income countries as most informal caregivers have been identified as relying on spiritual and religious coping [102,103]. Recommended strategies such as the provision of financial and social support as well as policy and legislature strategies call for an urgent need for policies in mental health to shift focus toward integrating caregiver-oriented services into practice. In addition, these findings highlight the need to adopt an intersectoral approach [104] in which various sectors such as religious and spiritual organizations provide mental health services to extend their accessibility to informal caregivers.

Limitations

In alignment with Arksey and O’Malley, this review did not include a critical appraisal to ascertain the quality of the studies as the purpose of this scoping review was to map existing strategies for alleviating caregiver burden and report on the outcomes as stated by the authors. In their study, Woo et al [105] cautioned that the exclusion of critical appraisals in scoping reviews means that the review cannot ascertain the research gaps that it aims to address if the included studies are of poor methodological quality. Although this scoping review provides evidence on the existing strategies for alleviating caregiver burden, it is important that the suggested strategies be evaluated in the specific context to ensure their effectiveness before implementation in clinical practice. This step was omitted as the purpose of this review was to map the available intervention strategies and the outcomes reported by the authors in alleviating burden among informal caregivers in low- and middle-income countries. This review only included studies conducted in low- and middle-income countries, and only papers written in English were considered, thus reducing the extent and scope of the evidence on the strategies for alleviating burden among informal caregivers regarding mental health.

Conclusions

The findings of this scoping review provided the authors with categories they can use to develop semistructured interview guides to use when exploring the existing formal and informal community mental health services to alleviate the burden of informal caregivers in rural South Africa. The categories outline the different types of strategies that can be used to alleviate caregiver burden, in particular the types of strategies offered to informal caregivers and the intervention developers and deliverers in and outside the mental health care system. Although most of the included studies (29/44, 66%) implemented these strategies, a few studies conducted in other low- and middle-income countries (15/44, 34%) recommended the use of these interventions to alleviate caregiver burden. Future studies from low- and middle-income countries in other continents, including Africa and South and North America, should address this gap in the research by evaluating these intervention strategies to alleviate the burden of informal caregivers.

Critical appraisal, which is used to ascertain the quality of the studies, was omitted as the purpose of this scoping review was to map existing strategies for alleviating caregiver burden and report on the outcomes as stated by the authors. Future studies should conduct quality appraisals to establish the effectiveness of these strategies in alleviating the burden of informal caregivers. Peer-facilitated support group interventions, although not frequently implemented, were identified as having long-term benefits compared with frequently implemented interventions such as psychoeducation. It is recommended that future research be directed at implementing and evaluating these interventions to alleviate burden in low- and middle-income countries.

Acknowledgments

The authors wish to acknowledge Mr Devind Peter, a librarian at the Wits Health Sciences Library, for assisting with the development of the search strategy and retrieval of the literature sources for this scoping review. This study was supported by the Consortium for Advanced Research Training in Africa. The Consortium for Advanced Research Training in Africa is jointly led by the African Population and Health Research Center and the University of the Witwatersand and funded by the Carnegie
Corporation of New York (grant G-19-57145); Swedish International Development Cooperation Agency (grant 54100113); Uppsala Monitoring Center; Norwegian Agency for Development Cooperation; and Wellcome Trust (reference 107768/Z/15/Z) and UK Foreign, Commonwealth, and Development Office with support from the Developing Excellence in Leadership, Training, and Science in Africa program. The statements made and views expressed are solely the responsibility of the fellow.

Data Availability
All data generated or analyzed during this study are included in this published paper.

Authors’ Contributions
OS completed the literature search. OS and DC completed the screening of titles, abstracts, and full texts of the identified studies. In addition, they extracted data from all the included studies and completed the draft of the manuscript. FA and NGN reviewed the manuscript draft. OS incorporated the feedback from the authors for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Data extraction templates.
[DOCX File, 23 KB - ijmr_v13i1e48587_app1.docx ]

Multimedia Appendix 2
PRISMA-ScR Checklist.
[PDF File (Adobe PDF File), 156 KB - ijmr_v13i1e48587_app2.pdf ]

References

https://www.i-jmr.org/2024/1/e48587 Interact J Med Res 2024 | vol. 13 | e48587 | p.59 (page number not for citation purposes)


103. Modise TP, Mokgaola IO, Sehularo LA. Coping mechanisms used by the families of mental health care users in Mahikeng sub-district, North West province. Health SA 2021 Aug 16;26:1586 [FREE Full text] [doi: 10.4102/hsag.v26i0.1586] [Medline: 34522392]


Abbreviations

CBT: cognitive behavioral therapy
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

©Olinidah Silaule, Daleen Casteleijn, Fasloen Adams, Nokuthula Gloria Nkosi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.01.2024. 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.
Review

Digital Methods for the Spiritual and Mental Health of Generation Z: Scoping Review

Susanna Y Park1*, PhD; Bridgette Do2*, MPH, PhD; Jacqlyn Yourell2, MS, PhD; Janice Hermer3, MA; Jennifer Huberty2, PhD

1Skylight, Radiant Foundation, Salt Lake City, UT, United States
2Fit Minded Inc, Phoenix, AZ, United States
3Arizona State University, Tempe, AZ, United States
*these authors contributed equally

Corresponding Author:
Susanna Y Park, PhD
Skylight
Radiant Foundation
55 N 300 W Suite 800
Salt Lake City, UT, 84101
United States
Phone: 1 (720) 675 7228
Email: susanna@skylight.org

Abstract

Background: Generation Z (Gen Z) includes individuals born between 1995 and 2012. These individuals experience high rates of anxiety and depression. Most Gen Z individuals identify with being spiritual, and aspects from religion and spirituality can be integrated into mental health treatment and care as both are related to lower levels of depression. However, research on the spiritual and mental health of Gen Z is sparse. To date, there are no systematic or scoping reviews on digital methods to address the spiritual and mental health of Gen Z.

Objective: This scoping review aimed to describe the current state of digital methods to address spiritual and mental health among Gen Z, identify the knowledge gaps, and make suggestions for how to leverage digital spiritual and mental health interventions for Gen Z.

Methods: A comprehensive literature search was conducted in PubMed, Scopus, PsycInfo, CINAHL, Education Full Text, Google Scholar, SocIndex, and Sociological Abstracts. The inclusion criteria were as follows: (1) study population born between 1995 and 2012 (ie, Gen Z); (2) reporting on spiritual health or well-being, spirituality or religion, and mental health or well-being; (3) reporting on using digital methods; (4) publication in 1996 or beyond; (5) human subject research; (6) full text availability in English; (7) primary research study design; and (8) peer-reviewed article. Two authors screened articles and subsequently extracted data from the included articles to describe the available evidence.

Results: A total of 413 articles were screened at the title and abstract levels, of which 27 were further assessed with full text for eligibility. Five studies met the inclusion criteria, and data were extracted to summarize study characteristics and findings. The studies were performed across 4 different countries. There were 2 mixed-methods studies (South Africa and Canada), 2 cross-sectional studies (China and United States), and 1 randomized controlled trial (United States). Of these studies, only 2 discussed digital interventions (a text messaging–based intervention to improve spiritual and mental health, and a feasibility study for a mental health app). Other studies had a digital component with minor or unclear spiritual and mental health measures. Overall, there was a lack of consistency in how spiritual and mental health were measured.

Conclusions: Few studies have focused on assessing the spiritual and mental health of Gen Z in the digital context, and no research to date has examined a digital spiritual and mental health application among Gen Z. Research is needed to inform the development and evaluation of approaches to address the spiritual and mental health of Gen Z via digital means (eg, mobile apps).

KEYWORDS
Generation Z; Gen Z; spiritual health; digital mental health; spirituality

Introduction

The digital age is marked by widespread internet use and the ability to quickly communicate and find information online. Individuals born from 1995 to 2012 are considered “digital natives,” as they are the first generation to live in an age where technology and the internet are accessible at all times [1]. The Pew Research Center considers these individuals as part of Generation Z (“Gen Z”) and identifies the beginning of this generation to be in 1997 [1]. For the purpose of our research, we used the definition of Gen Z by Katz et al [2], who defined Gen Z as individuals born from 1995 to 2012, to account for those born when the World Wide Web made its public debut in 1995.

As a generation that grew up with technology, digital devices are familiar and seamlessly woven into the daily routines of Gen Z. It has been reported that 95% of Gen Z individuals have access to a smartphone [3]. In a study of 1000 Gen Z individuals aged 13 to 25 years, more than half spent 4 or more hours online compared with just 28% of all US adults who spent 4 or more hours online [4]. In a world where it is nearly impossible to socialize, work, and get an education without technology, Gen Z individuals are “always on,” and this is associated with higher rates of depression, attention deficit disorder or attention deficit hyperactivity disorder, and technology addiction [5]. Compared with other generations, Gen Z individuals spend more time alone or on digital communication platforms than engaging in in-person interactions. Between 2000 and 2015, Gen Z high school seniors spent an hour less on in-person social interactions compared with early millennials [6]. Adolescents who spend more time on social media than in in-person interactions are the loneliest compared with those who spend less time on social media. Moreover, between 2007 and 2018, there were great increases in the relative percentage rates for self-injury (47%), seriously considering suicide (76%), and suicide attempts (58%) among Gen Z [7].

The American Psychological Association (APA) states that Gen Z individuals are more likely to report mental health concerns (eg, depression and anxiety) than previous generations [8]. In a 2022 survey of 1055 Gen Z adults, 1 out of 4 reported having more bad days than good within a 1-month time frame. More than 2 out of 5 (42%) had a diagnosed mental health condition, with more than a quarter of those being diagnosed during the COVID-19 pandemic (March 2020) or later. Anxiety and depression are the 2 largest mental health issues among Gen Z, with 9 out of 10 individuals diagnosed with a mental health condition having anxiety and 8 out of 10 having depression [3]. Notably, Gen Z individuals are the most comfortable discussing their mental health [3,8]. One third of Gen Z individuals report posting about their mental health on social media. They also attend therapy and are willing to pay out of pocket for mental health care and services [3]. Despite the comfort of Gen Z in talking about their mental health, there is a crucial need to address the high rates of anxiety, depression, and other health issues that they experience [8].

Spirituality may be an untapped resource to address the mental health crisis experienced by Gen Z today. While spirituality can serve as a component within organized religion, the 2 aspects are distinct. Religion is an organized belief or specific set of practices focusing on a higher power (ie, Christian, Muslim, Buddhist, etc) [9]. Spirituality is a broader concept in which individuals seek connection to self, others, nature, and a sacred or higher being [10]. Individuals may identify with being either religious or spiritual, or both. Gen Z individuals do not necessarily identify with a particular religion or belief but instead practice spirituality. Only half of Gen Z individuals report turning to their faith for support in times of uncertainty [11], and they are more likely to engage in spiritual practices than religious practices [12]. In a study of 10,000 Gen Z individuals aged 13 to 25 years, 68% considered themselves religious and 77% considered themselves spiritual [13]. Gen Z individuals define spirituality as autonomous and faith unbundled, and it is inclusive of all faiths and practices [11,13].

Spirituality is related to several positive health and psychosocial outcomes, namely greater mental health [14]. A recently updated review of the literature on the relationship between spirituality and mental health found that greater spirituality was associated with lower depressive symptoms, lower suicidality, and lower substance abuse [14]. Gen Z individuals face some of the highest rates of mental health conditions (eg, depression) [15,16]; thus, spirituality should be considered in addressing youth mental health today. In the aforementioned report by Singer [13], the majority of Gen Z individuals attributed their spiritual connection to their positive mental health state. Another aspect of mental health that is influenced by spirituality is quality of life among chronically and terminally ill patients. Palliative care patients who struggle with spirituality report poorer quality of life compared with those who feel stable with their spirituality [17]. Additionally, teens and young adults with cancer mention searching for meaning, hope, and life perspectives, even though they may not consider themselves as spiritual [18]. Interventions that promote spiritual well-being (one’s sense of purpose, meaning in life, and connection to something greater [19]) may be a powerful resource for improving mental health in Gen Z.

Research on digital mental health interventions and spirituality exist separately. Little is known about digital methods (eg, mobile apps, text messaging, etc) that incorporate both spiritual and mental health among Gen Z. Scoping or systematic reviews on this topic are nonexistent, and research on this topic is very limited. Given that technology is woven into the daily lives of Gen Z, digital mobile apps that promote spirituality may offer a novel approach to supporting the mental health of Gen Z adolescents and young adults. Therefore, the purpose of this scoping review was to describe the current state of digital methods to address spiritual and mental health among Gen Z, identify the knowledge gaps, and make suggestions for how to leverage digital spiritual and mental health interventions for Gen Z.
Methods

Eligibility Criteria
The inclusion criteria for targeted articles were as follows: (1) study population born between 1995 and 2012 (ie, Gen Z); (2) reporting on spiritual health or well-being, spirituality or religion, and mental health or well-being; (3) reporting on using digital methods; (4) publication in 1996 or beyond; (5) human subject research; (6) full text availability in English; (7) primary research study design; and (8) peer-reviewed article.

Information Sources
Guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping reviews (PRISMA-ScR) [20], searches were conducted in PubMed, Scopus, PsycINFO (EBSCO), CINAHL Plus with Full Text (ProQuest), Education Full Text (H.W. Wilson), Google Scholar, SocIndex with Full Text, and Sociological Abstracts (ProQuest) on May 2 and 3, 2023, by a health sciences librarian (J. Hermer) at Arizona State University. The same information sources were searched on September 21-26, 2023, and again on November 7-8, 2023, and were also included in the results.ERIC and Atla Religion Database were also searched, but they returned no results.

Search Strategy
The searches were optimized for each individual database but included a combination of keywords and subject headings for the following 4 categories: Generation Z or Gen Z; spirituality or religion; mental well-being; and mobile health, mHealth, and eHealth (Multimedia Appendix 1). Owing to the very limited nature of the results, only language filters were applied to ensure that all relevant literature was available to be screened. All records were imported into Zotero to check if any articles were retracted and then into Covidence systematic review software for deduplication and screening [21]. Apart from the initial search, 2 additional searches were conducted to ensure that all relevant literature was found and to add any recent literature from the following 6 months. A search of all databases was performed on September 21-26, 2023, using the additional keywords of “faith,” “transcendence,” and “life purpose” or “existential needs,” and a final search was performed on November 7, 2023, using the additional keywords of “spiritual” and “religion.” The scoping review protocol was registered online in Open Science Framework (OSF) on May 16, 2023.

Selection Process
The selection process was completed entirely on Covidence. Prior to reviewing titles and abstracts, duplicate titles were eliminated by Covidence. Two authors (SYP and BD) screened all titles and abstracts independently and were blinded to each other’s decisions. Any disagreements were discussed between the authors SYP and BD, and agreed upon before full-text review. Agreement scores for article selection between the 2 authors were not logged in Covidence for the initial search; however, the agreement scores were 85% (81/95) and 90% (80/87) for reviewing titles and abstracts between the 2 authors for the September and November searches, respectively. For full-text article review, the author SYP independently reviewed half of the articles, and the author BD independently reviewed the remaining articles. The authors SYP and BD deliberated with each other, and with J Huberty and JY if there were any questions regarding inclusion based on the article eligibility criteria. Articles were excluded if the study population included Gen Z but did not explicitly distinguish Gen Z in the population sample and the results were not disaggregated by age. The final articles included in the review were agreed upon by all 4 authors.

Data Collection Process and Synthesis Methods
Prior to the search, all authors agreed on the following characteristics for data extraction and synthesis: title; authors; study country; study objectives; study design; data collection timeframe; recruitment methods; sample size; participant characteristics; description of digital methods; constructs related to religion, spirituality, or spiritual well-being; assessment or measure for religion, spirituality, or spiritual well-being; constructs related to mental health; assessment or measure for mental health; main findings; and study limitations. These characteristics were selected to ensure a detailed understanding of available literature as it relates to the goal of the scoping review. For articles included in the scoping review, the author SYP independently extracted data based on the a priori characteristics for half of the articles and the author BD independently extracted data from the remaining articles. Given that each article had data extracted by a single author, there were no agreement scores. After evaluating the included articles, study characteristics and main findings were summarized in a descriptive manner.

Some studies that initially appeared eligible for this review were ultimately excluded because they did not meet specific criteria. For instance, a cross-sectional study on the perceptions of 475 Gen Z individuals and young millennials and their use of a spiritual self-care app [22] was originally included when reviewing titles and abstracts. However, upon full-text review, we found that the study results did not distinguish between Gen Z individuals and young millennials, thus failing to meet our review’s criteria (ie, Gen Z only).

Results

Study Characteristics
Details on article selection are illustrated in a PRISMA diagram (Figure 1). Our search identified 824 articles from 8 databases based on the search terms. After removing 411 duplicates in Covidence, 413 articles were screened by title and abstract. After the initial screening based on the inclusion criteria, the full texts of the remaining 27 articles were screened. Review articles were not included in the final review, but references were screened to see if any additional literature was admissible. Ultimately, the scoping review included 5 articles. Of these 5 articles, 1 [23] was included in the review from the updated search conducted in September 2023 and 1 [24] was included from the updated search in November 2023. Characteristics and results of the studies are summarized in Table 1.
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of article selection.

- **Identification**
  - Records identified from databases (n=824)
    - PubMed (n=165)
    - Scopus (n=57)
    - PsycInfo (n=63)
    - CINAHL (n=28)
    - Google Scholar (n=190)
    - SociNEX (n=11)
    - Sociological Abstracts (n=300)
    - Education Full Text (n=9)
  - Duplicate records removed (n=411)

- **Screening**
  - Titles and abstracts screened (n=413)
  - Records excluded (n=386)
    - Articles not retrieved (full text unavailable) (n=1)
  - Full-text articles sought for retrieval (n=28)
  - Full-text articles assessed for eligibility (n=27)
    - Articles excluded:
      - Wrong study population (ie, not Generation Z) (n=10)
      - Did not examine spirituality/religion (n=2)
      - No digital methods (n=1)
      - Wrong study design (eg, review) (n=6)
      - Not peer-reviewed (n=3)

- **Included**
  - Articles included in scoping review (n=5)
Table 1. Characteristics and main findings of the included studies.

<table>
<thead>
<tr>
<th>Reference and study country</th>
<th>Study design</th>
<th>Sample characteristics</th>
<th>Digital aspect</th>
<th>Religion/spirituality measures</th>
<th>Mental health measures</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mindu et al [25], 2023 (South Africa)</td>
<td>Mixed-methods study</td>
<td>93 youth and young adults aged 16-24 years; 44% female</td>
<td>Assessed participants’ knowledge and preferences for a digital mental health app</td>
<td>Where do youth seek treatment or assistance when they have mental health problems (Response options: Visit a spiritual healer; Go to a church for prayers; Traditional medicine; Clinic/hospital; Visit a health care worker)</td>
<td>Mental health awareness (eg, taught about mental health and no prior education on mental health) Mental health conditions affecting youth (eg, substance abuse or misuse, posttraumatic stress disorder, depression, and anger)</td>
<td>• No participants had experience using a mental health app, but 99% indicated mental apps are important and can benefit youth. • Religious and cultural beliefs were a barrier to using digital platforms (eg, social media).</td>
</tr>
<tr>
<td>Gao et al [26], 2021 (China)</td>
<td>Cross-sectional study</td>
<td>1017 first-year college students (mean age 19 years); 77.8% female</td>
<td>Participants were recruited through an eHealth application to complete a survey</td>
<td>Religion (Response options: No religion; Buddhism; Christian; Others)</td>
<td>Generalized Anxiety Disorder-7 (GAD-7) and depression; Patient Health Questionnaire-9 (PHQ-9) [27,28]</td>
<td>• 95.3% indicated having no religion. • Belief in Christianity and in Buddhism were associated with greater anxiety.</td>
</tr>
<tr>
<td>Craig Rushing et al [29], 2021 (United States)</td>
<td>Randomized controlled trial</td>
<td>833 American Indian or Alaska Native teenagers and young adults aged 15-24 years; 66.3% female</td>
<td>The 2 intervention arms included 3 SMS text messages per week for 8 weeks with information, role model videos, images, and engagement opportunities (eg, reply for more information, resource links, etc)</td>
<td>Rate your spiritual health (Response options: Excellent; Very good; Good; Fair)</td>
<td>Rate your mental health (Response options: Excellent; Very good; Good; Fair)</td>
<td>• No significant differences between the 2 intervention arms within subjects. • Mean scores of perceived health (physical, mental, and spiritual) significantly increased over time for both intervention arms. • Those who reported better health also reported greater cultural resilience, identity, and cultural pride.</td>
</tr>
<tr>
<td>Reed et al [23], 2022 (United States)</td>
<td>Cross-sectional study</td>
<td>349 American Indian or Alaska Native youth aged 15-24 years; 71.1% female</td>
<td>Assessed participants’ use of media technologies (ie, media types, frequency, and duration) and how they use media technologies (ie, online behaviors and activities)</td>
<td>Select the top 3 health topics from a list of 15, including spiritual health</td>
<td>Self-reported mental health (How good is your mental health? Response options: 4-point Likert-type scale; 4=Excellent and 1=Poor) Select the top 3 health topics from a list of 15, including mental health</td>
<td>• 53.5% of participants relied heavily on the internet to access health information. • Nonsexual and gender minority participants reported better mental health than sexual and gender minority participants. • The top 3 most important health topics were Native identity, mental health, and social justice and equality. • Spiritual health was selected as the most important health topic 14% of the time.</td>
</tr>
</tbody>
</table>
Study Design and Study and Sample Characteristics

All 5 studies examined Gen Z (adolescents and young adults; aged 11-24 years; born between 1995 and 2012). Of the 5 articles, 2 were published in 2023, 1 was published in 2022, and 2 were published in 2021. There were 2 mixed-methods studies (1 in South Africa [25] and 1 in Canada [24]), 2 cross-sectional studies (1 in China [26] and 1 in the United States [23]), and 1 randomized controlled trial in the United States [29]. Sample sizes varied across the 5 studies, ranging from 5 to 1017.

Study by Mindu et al, 2023

Mindu et al [25] conducted a mixed-methods study in South Africa to inform the implementation of a mobile phone–based mental health intervention [25]. The study did not use any specified measures for mental health but asked questions via quantitative and qualitative methods to understand participants’ understandings or perceptions of mental health. Questions addressed (1) mental health awareness; (2) digital interventions for mental health; (3) access to digital devices and the internet; (4) preferences for digital mental health interventions; and (5) barriers to the use of digital mental health innovations. Quantitative results showed that almost half (49%) of the participants had heard about mental health apps, none had experience using one, and 99% indicated that mental health apps are important and can benefit youth. These results underscore the severe lack of mobile mental health apps designed for individuals in South Africa. Qualitative results revealed that religious and cultural beliefs were a barrier to using digital platforms (eg, social media), and the authors highlighted the crucial need to develop culturally appropriate and relevant digital apps that represent those who they serve. Overall, participants expressed high interest in using a digital mental health app (eg, social media) to learn about mental health and seek resources. The study authors acknowledged that while no digital methods or interventions were implemented in the study, the results provided a unique Gen Z perspective on the usability of digital methods.

Study by Gao et al, 2021

Gao et al [26] conducted a cross-sectional study in China among first-year Chinese college students. Using a health management app (Residents e-Health), a questionnaire was distributed to examine depression and anxiety, and their associations with other health-related constructs, such as stress and nutrition. The authors examined participants’ religious beliefs (ie, Christian, Buddhism, other, and no religion) as a correlate of anxiety and depression. However, most of the sample (95%) did not have a religion. Belief in Christianity and belief in Buddhism were associated with greater anxiety in the sample. The study authors noted that the large number of participants ascribing to no religion limited the understanding of the relationship between religion and mental health. The overall conclusion of the study indicated that early lifestyle interventions assessing religion, as well as other demographic and behavioral factors, are important for understanding the factors contributing to mental health in Gen Z.

Study by Craig Rushing et al, 2021

Craig Rushing et al [29] conducted a randomized controlled trial in the United States to examine the efficacy of an mHealth intervention (BRAVE) for physical, mental, and spiritual health: resilience; self-esteem; and coping and help-seeking skills among American Indian or Alaska Native teenagers and young adults. Participants were randomized to participate in 1 of 2 groups: (1) An 8-week intervention arm involving 3 SMS text messages per week highlighting common coping strategies, preferred wellness strategies, help-seeking skills, and related protective factors such as cultural resilience, identity, and cultural pride; or (2) An 8-week control arm involving 3 SMS text messages per week designed to elevate and reaffirm Native voices in STEM (science, technology, engineering, and mathematics) and medicine. Participants were in both arms, and both had messaging that included a combination of information,
role model videos, images, and opportunities of engagement (eg, reply for more information and links to access resources). The findings indicated that there were no significant differences between the 2 study arms within subjects, such that participants in the intervention arm did not report better outcomes than those in the control arm. However, mean scores of perceived health (ie, physical, mental, and spiritual) significantly increased over time in both arms. In addition, participants who reported better health also reported greater cultural resilience, identity, and cultural pride. The percentage of participants who used the resources and information in the intervention arm text messages also increased over 5 months. The study authors acknowledged some limitations, including high or favorable survey measure outcomes at baseline, only a 1-week break between receiving interventions in the arms, and control messaging likely being novel and helpful. Another limitation to note is the assessment of perceived health, which combined 3 separate survey measures assessing perceived physical, mental, and spiritual health. The combination of 3 different aspects of health makes it challenging to derive valid inferences regarding mental or spiritual health on their own. Overall, the intervention demonstrated improved health outcomes and underscores the acceptability of text messaging to promote and support well-being.

**Study by Reed et al, 2022**

Reed et al [23] conducted a cross-sectional study in the United States, in which 349 American Indian or Alaska Native youth (aged 15-24 years) were asked about the extent to which they use media technologies, how they use technologies, and their health priorities. Several trends were revealed. The majority of participants (64.7%) reported sending 1 to 50 text messages a day. Instagram was the most popular daily technology used, and 65.3% of participants reported using social media 3-7 hours per day. Participants also self-rated their mental health. The findings indicated that nonsexual and gender minority youth (56.7%) reported better mental health than sexual and gender minority youth (36.4%). To better understand important health topics, participants were asked to select their top 3 health topics from a list. The most popular topic selected was Native identity or cultural pride (73%), followed by mental health (57%) and social justice and equality (31%). Spiritual health was selected by 14% of youth in their top 3 health topics. While spiritual health was not among the top health topics selected, it is important to note that the range of options offered spanned across several categories (eg, social justice or inequality, alcohol and drug use or abuse, and the environment). Overall, the study authors concluded that building resources that foster cultural pride and positive identity must be included in any programs or technologies for addressing mental health among American Indian or Alaska Native youth.

**Study by Au-Yeung et al, 2023**

Au-Yeung et al [24] conducted a Haudenosaunee (Canada’s largest First Nations reserve, Six Nations of the Grand River) community-based study in which 5 Haudenosaunee youth (aged 11-16 years) tested the JoyPop mobile app (available in English and French) that is designed to promote resilience among youth. The app offers breathing exercises, mood tracking, journaling, personalized social support, a 24-hour helpline, and games. The results indicated that participants had mixed ratings on their self-reported mental wellness, ranging from “fair” to “very good,” and that they used the app 1 to 3 times a day. Of the 5 participants, 4 were interviewed about their experiences using the app, perspectives of mental wellness, and characteristics of a “Good Mind” (an Indigenous concept on the physical, psychological, and spiritual journey that maintains balance and harmony in a person). Interview participants reported that the app was easy to use and esthetically pleasing. They also enjoyed all of the app’s features, with the exception of the Circle of Trust feature (ie, personalizable social support). Interview participants identified positivity and happiness, understanding emotions, acts of kindness, personal hobbies, and positive body language as important to their mental wellness. Participants also discussed important characteristics of a Good Mind, such as positivity, kindness, and connecting with nature. Overall, the app was favorable to the participants, but they suggested incorporating specific features like words in their own language and Indigenous visuals (eg, feathers and clan animals). While the app lacked explicit content on a Good Mind, the authors recommended incorporating concepts of a Good Mind to enhance its relatability to Haudenosaunee youth, given its cultural significance to the Haudenosaunee people. Although the study had a small sample size, the authors concluded that mobile health interventions can be beneficial to Indigenous youth, as mental health apps continue to be of interest and Indigenous cultures value the promotion of health and resilience. Further, Indigenous tribes across North America have unique perspectives, and pan-Indigenous resilience apps like JoyPop will need to be tailored to specific cultural contexts.

**Discussion**

**Overview**

The purpose of this scoping review was to describe the current state of digital methods to address spiritual and mental health among Gen Z, identify the knowledge gaps, and make suggestions for how to leverage digital spiritual and mental health interventions for Gen Z. A comprehensive literature search across 8 databases identified only 5 relevant studies, emphasizing the significant lack of published research on digital methods to address the spiritual and mental health of Gen Z. Among the 5 studies, only 2 discussed digital interventions, and of these studies, 1 examined a text messaging–based intervention to improve spiritual and mental health and reported improvements in spiritual and mental health over time [27], and 1 examined the feasibility of a mental health app [24]. The sparse available literature limits conclusions on the current state of digital methods to address spiritual and mental health, and warrants future research to address these gaps.

**Current State of Digital Methods for Spiritual and Mental Health**

Gen Z individuals are facing a mental health crisis as they experience high rates of depression and anxiety [8,16]. For example, one of the included studies reported that the prevalences of anxiety and depression among college freshmen were 40.3% and 45.3%, respectively [26]. Gen Z individuals also report feeling lonely, having low self-confidence, and being...
distressed about the future [3,5,12]. All 5 studies included in this review discussed the overall well-being of Gen Z and, to some extent, aspects of spirituality. The findings suggest that spirituality may play a role in the mental health of Gen Z and should be considered in the development and implementation of future digital applications that address mental health. The study by Mindu et al [25] particularly underscores the need for customizability in mental health applications considering that participants expressed that using digital platforms (ie, social media) conflicted with their religious and cultural beliefs. Thus, the introduction of applications that allow users to engage in practices and view content that aligns with their values and beliefs is a potential avenue for combating this barrier and could in turn strengthen one’s spirituality. However, research is needed to determine the acceptability of digital mental health applications that involve spiritual content.

Nearly every facet of the lives of Gen Z involves technology, for example, using computers for school or work, using mobile apps to order food, and using social media and texting to communicate with friends. Despite the widespread use of technology among Gen Z and the increasing number of research studies employing digital methods to test and deliver mental health programs and interventions [31], only 1 of the studies included in this review examined the effects of a digital intervention on spiritual and mental health [27] and I assessed the feasibility of a mental health mobile app [24]. The lack of digital methods or interventions to address the spiritual and mental health of Gen Z warrants the development of accessible ways for Gen Z to practice spiritual and mental self-care. Gen Z individuals spend more time (4 or more hours daily) on social media compared with other generations [3], and poor mental health is often attributed to social media use [32]. However, social media is not necessarily harmful and is instead dependent upon what Gen Z individuals do and see online, their pre-existing strengths or vulnerabilities, and the environment in which they are raised (eg, parental monitoring) [32,33]. Digital methods for spiritual wellness that are specifically targeted for Gen Z and built to empower Gen Z to practice self-care and build healthy coping mechanisms may benefit the mental health of Gen Z. Results from this review illustrate the acceptability of a mobile intervention promoting spiritual and mental health [24,29], and the interest Gen Z individuals have toward using digital methods to address their spiritual and mental health [25]. However, given that little is known about this topic, more research is needed to truly grasp the feasibility and efficacy of digital approaches to address spiritual and mental health with regard to the well-being of Gen Z.

Gaps in the Literature
Published literature or research about the spiritual and mental health of Gen Z is limited, reporting mostly mental health statistics rather than examining determinants of mental health or interventions for improving mental health (eg, spirituality) [1,33]. Research to date on spirituality in Gen Z has only been performed by faith-based organizations or by nonacademic research institutions and has mainly focused on comparing the views of Gen Z on religion and spirituality to the views of other generations [11,12,34]. For example, the majority of Gen Z individuals (77%) identify as spiritual, preferring to ascribe to a set of values from various beliefs [13]. While we know that the majority of Gen Z individuals identify as spiritual, empirical research that specifically examines their spiritual practices and preferences, overall spiritual well-being, and associations with mental and physical health outcomes is warranted. Among the 5 studies included in this review, 3 examined spiritual health. The randomized controlled trial [29] asked participants to rate their spiritual health using a combined measure for overall health that included physical, mental, and spiritual health, limiting the ability to assess spiritual well-being explicitly among these participants. The cross-sectional survey [23] asked participants to select the top 3 health topics important to them (eg, mental health, spiritual health, and Native identity), but the study did not examine any associations between spiritual health and mental health. Spiritual health was not defined for participants in either study. The mixed-methods study in Canada [24] interviewed participants who used an app designed to promote resilience in youth with regard to a Good Mind, an Indigenous concept on the physical, psychological, and spiritual journey that maintains balance and harmony in a person. However, this concept is specific to the Haudenosaunee people; therefore, the findings may not be applicable to the perspectives on spirituality of other young communities. In addition, only 1 article administered a valid measure to assess mental health, spirituality, or spiritual well-being (ie, Generalized Anxiety Disorder-7 [GAD-7]) [26]. The validity and reliability of instruments used to measure mental health and spirituality or spiritual well-being in the study samples were not reported in any of the included articles. Identified gaps offer research opportunities to comprehensively examine spiritual and mental health among Gen Z.

Suggestions for Leveraging Digital Spiritual and Mental Health Interventions for Gen Z
Gen Z individuals are remarkably familiar with navigating digital spaces and integrating spirituality into their lives, and are the most comfortable talking about their mental health compared with other generations [3]. Gen Z individuals who have a spiritual connection have better perceptions about their mental health and believe their spiritual health contributes to their mental health, compared with those who do not have a spiritual connection [11]. Since Gen Z individuals primarily consume information through technology, digital interventions are a promising method to teach and facilitate various practices in spiritual and mental self-care. For example, the digital intervention content of Craig Rushing et al [29] included wellness strategies, such as self-care and goal setting, which resulted in improved perceived health (eg, physical, mental, and spiritual) for participants. In the study by Mindu et al [25], digital mobile interventions through social media were concluded to be potentially useful to increase mental health literacy and knowledge of resources. Additionally, the study by Gao et al [26] suggested that early interventions that target the lifestyle behaviors of Gen Z (ie, smoking) can improve their depression and anxiety. A spiritual self-care app tailored for Gen Z, for example, may allow Gen Z to engage with full autonomy and convenience. Rather than requiring Gen Z to seek out places to practice religious or spiritual beliefs, a spiritual self-care app can be available wherever they are and
whenever they want. Digital spiritual self-care interventions should also consider incorporating topics that are deemed important to Gen Z, such as cultural relevance, inclusivity, social justice, and nature [23,24].

Mobile apps grant users the ability to customize their experience, which vastly differs from traditional means of practicing religion that ascribe to a predetermined set of beliefs, values, and practices that are often fixed [13]. Access to technology enables Gen Z individuals to autonomously decide which aspects of different spiritual beliefs and practices they resonate with. Most digital apps include features that allow users to customize their experience and the content they engage with. Thus, a mental health app that includes various components of spiritual practices that users can choose from may provide a new way for young people to tend to their spiritual and mental health. Based on this review, research examining the feasibility, acceptability, and efficacy of digital health tools specifically targeting spiritual and mental health among Gen Z is absent, indicating the need for research in this area. Researchers, companies, and nonprofit organizations can leverage existing digital spaces that Gen Z individuals frequently use (eg, Instagram and TikTok) to garner feedback on what they might desire in a digital spiritual and mental health intervention. For example, 2 of the studies [24,29] incorporated feedback from Indigenous youth on existing digital methods (eg, SMS text messaging and JoyPop) to inform the development of their digital content. Spiritual well-being apps, for example, can be used to target Gen Z and deliver evidence-based content that integrates spirituality and mental health. These studies also emphasize the importance of culturally relevant interventions that can speak to diverse cultural backgrounds and beliefs among Gen Z.

Strengths and Limitations of This Review

There are multiple strengths of this review. First, this scoping review indicates that there is a limited knowledge base surrounding digital methods for addressing the spiritual and mental health of Gen Z and summarizes the current state of the literature on this topic. This is the first scoping review to address this topic and highlights a crucial gap in supporting young people’s mental health. Additionally, a librarian was consulted and involved in the search process to bolster the rigor and accuracy of the review. Covidence was used to minimize human error in screening eligible articles. Along with these strengths, there are some limitations in this review. First, a scoping review limits the objective understanding of a topic such that quantitative results cannot be compiled to determine effect sizes across studies. Second, database searches are not uniform and require nuanced search methodologies, which can result in relevant studies being missed. Third, manuscripts during the full-text screening stage were split between the authors SYP and BD, and thus, manuscripts were assessed by a single rater. Finally, research on this topic is severely lacking, which limits the number of articles included in this review and impacts the ability to construct a cohesive narrative or draw definitive conclusions about the state of the field. This underscores the necessity of this review to highlight the gaps and urge further investigation.

Scientific Contribution

The goal of this scoping review was to assess the extent to which research on digital methods to address the mental and spiritual health of Gen Z has been conducted. The scoping review revealed a lack of available research on spirituality and mental health. Specifically, there is a dearth of studies on the use of digital methods to deliver spiritual well-being for mental health in Gen Z. Despite an increase in mental health concerns among Gen Z and the growing body of evidence on the beneficial effects of spiritual self-care on mental health, few published articles touch on this topic. The findings from this review highlight the opportunity for addressing the mental and spiritual health of Gen Z through digital methods (eg, mobile apps). The use of digital methods to address mental health is a growing area of research; however, spirituality and spiritual self-care have received little attention. There is potential for researchers to examine spiritual self-care, which is delivered through digital methods, and its impact on populations experiencing significant mental health problems. Overall, the scoping review underscores the need for future research to examine the acceptability and feasibility of digital approaches to address spiritual and mental health among Gen Z.

Conclusion

This scoping review underscores the dearth of research surrounding digital methods to address spiritual and mental health among Gen Z. Considering that digital methods to address aspects of mental health are increasingly popular and effective [35,36], research is needed to examine digital platforms that address spiritual and mental health. This is especially pertinent for Gen Z individuals as they have some of the greatest rates of mental health issues and are the most digitally savvy generation to date, and most indicate that they are spiritual. Leveraging spirituality as a way to address mental health among Gen Z via digital means offers a novel and relevant approach for addressing the mental health crisis impacting young people today.

Acknowledgments

The authors would like to thank Clare Beatty for the organization of this project and Justin Jones for inspiring us to write this paper.

Data Availability

The full review process, including the search terms and databases used, is available as a supplementary file.
Authors' Contributions
SYP and BD wrote and reviewed this paper. J Hermer searched and curated articles for review. JY reviewed and edited the manuscript. J Huberty supervised the review.

Conflicts of Interest
SYP is currently the Behavioral Research Scientist for the Radiant Foundation and does not receive incentives for the outcomes of the research. J Huberty is the Chief Science Officer for the Radiant Foundation’s Skylight app. J Huberty was hired by the Radiant Foundation to lead scientific strategy and mentor the scientists for Skylight, but does not receive incentives for the outcomes of the research. BD and JY are employees of J Huberty and are independent from the Radiant Foundation leadership. The role of SYP and J Huberty is to ensure the quality of science regarding the Skylight app. J Hermer is a librarian at Arizona State University and is not associated with the Radiant Foundation or with J Huberty. The authors do not have stock in Skylight and receive no financial incentives from the sales of Skylight. The authors do not receive any financial benefit from the outcomes of this study.

Multimedia Appendix 1
Databases and search terms used for review.
[DOCX File, 23 KB - ijmr_v13i1e48929_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 239 KB - ijmr_v13i1e48929_app2.pdf]

References

https://www.i-jmr.org/2024/1/e48929 Interact J Med Res 2024 | vol. 13 | e48929 | p.74 (page number not for citation purposes)


**Abbreviations**

- **Gen Z**: Generation Z
- **PRISMA**: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Improvements in Neoplasm Classification in the International Classification of Diseases, Eleventh Revision: Systematic Comparative Study With the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision

Yicong Xu1*, MPH; Jingya Zhou2,3*, MPH; Hongxia Li4, MD; Dong Cai1, BA; Huanbing Zhu5*, MD, PhD; Shengdong Pan6*, MPA, LLM

1Medical Records Room, Department of Medical Administration, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
2Department of Medical Records, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
3Collaborating Center for the WHO Family of International Classifications in China, Beijing, China
4Department of Gastroenterology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
5Department of Hepatobiliary and Pancreatic Surgery, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
6Department of Medical Administration, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

*these authors contributed equally

Corresponding Author:
Shengdong Pan, MPA, LLM
Department of Medical Administration
The Second Affiliated Hospital, Zhejiang University School of Medicine
88 Jiefang Road
Hangzhou, 310009
China
Phone: 86 1 366 666 9123
Email: psd618@zju.edu.cn

Abstract

Background: The International Classification of Diseases, Eleventh Revision (ICD-11) improved neoplasm classification.

Objective: We aimed to study the alterations in the ICD-11 compared to the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (ICD-10-CCM) for neoplasm classification and to provide evidence supporting the transition to the ICD-11.

Methods: We downloaded public data files from the World Health Organization and the National Health Commission of the People’s Republic of China. The ICD-10-CCM neoplasm codes were manually recoded with the ICD-11 coding tool, and an ICD-10-CCM/ICD-11 mapping table was generated. The existing files and the ICD-10-CCM/ICD-11 mapping table were used to compare the coding, classification, and expression features of neoplasms between the ICD-10-CCM and ICD-11.

Results: The ICD-11 coding structure for neoplasms has dramatically changed. It provides advantages in coding granularity, coding capacity, and expression flexibility. In total, 27.4% (207/755) of ICD-10 codes and 38% (1359/3576) of ICD-10-CCM codes underwent grouping changes, which was a significantly different change ($\chi^2=30.3; P<.001$). Notably, 67.8% (2424/3576) of ICD-10-CCM codes could be fully represented by ICD-11 codes. Another 7% (252/3576) could be fully described by uniform resource identifiers. The ICD-11 had a significant difference in expression ability among the 4 ICD-10-CCM groups ($\chi^2=93.7; P<.001$), as well as a considerable difference between the changed and unchanged groups ($\chi^2=74.7; P<.001$). Expression ability negatively correlated with grouping changes ($r=-.144; P<.001$). In the ICD-10-CCM/ICD-11 mapping table, 60.5% (2164/3576) of codes were postcoordinated. The top 3 postcoordinated results were specific anatomy (1907/3576, 53.3%), histopathology (201/3576, 5.6%), and alternative severity 2 (70/3576, 2%). The expression ability of postcoordination was not fully reflected.
Conclusions: The ICD-11 includes many improvements in neoplasm classification, especially the new coding system, improved expression ability, and good semantic interoperability. The transition to the ICD-11 will inevitably bring challenges for clinicians, coders, policy makers and IT technicians, and many preparations will be necessary.

(KEYWORDS)
Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision; ICD-10; ICD-10-CCM; ICD-11; improvement; International Classification of Diseases, Eleventh Revision; International Classification of Diseases, Tenth Revision; International Classification of Diseases; neoplasm; transition

Introduction
The World Health Organization (WHO) adopted the International Classification of Diseases, Tenth Revision (ICD-10) in May 1990 [1]. The ICD-10 has been widely used in over 120 countries over the past 30 years. In multiple countries, expansions of the ICD-10, such as the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (ICD-10-CCM), based on the second edition [2], have been developed to meet specific requirements. The ICD-10-CCM has been used for national performance assessment and medical insurance payment in public hospitals. Given the advances in medical knowledge and health information, the WHO started working on the International Classification of Diseases, Eleventh Revision (ICD-11) in 2007 [3], and it was expected to be implemented by WHO member countries starting in January 2022.

The most critical changes in the ICD-11 were the establishment of a semantic knowledge base and the reconstruction of the coding system. It introduced the Foundation Component, the Common Ontology, and linearization, as well as the new concepts of precoordination and postcoordination [3-5]. The chapter on neoplasms has also changed. It adds crucial morphology to precoordination, having the highest proportion (98.9%) and the most dimensions (3.5) of postcoordination. Cancer has been one of the top 3 causes of death in China since 2005 [6]; thus, the coding system of the ICD-11 will significantly impact cause of death reporting, cancer registration, and disease diagnosis records. Stakeholders need to have a good understanding of the classification of neoplasms in the ICD-11.

This study analyzes the changes between the ICD-10-CCM and ICD-11 in terms of coding features, classification features, and expression features in neoplasm classification and hopes to provide evidence supporting the transition in China.

Materials
The following 3 existing, publicly available files were used in this paper: (1) ICD-10/ICD-11 mapping tables [7], (2) ICD-11 simple tabulation [8], and (3) the second revision of the ICD-10-CCM [9].

The first 2 files were downloaded from the WHO website, and the ICD-10-CCM was released by the National Health Commission of the People’s Republic of China.

Research Methods
The ICD-10-CCM neoplasm codes were manually recoded by the ICD-11 coding tool [10] based on the International Classification of Diseases, Eleventh Revision, Mortality and Morbidity Statistics (ICD-11 MMS) codes to generate an ICD-10-CCM/ICD-11 mapping table for neoplasms. The 3 existing files and the ICD-10-CCM/ICD-11 mapping table were used to analyze ICD-11 features, namely, the coding structure, coding capacity, grouping changes, expression ability, expression flexibility, and the expression of postcoordination in the neoplasm classification.

Mapping ICD-10-CCM to ICD-11 Codes
Due to the homology of the morphology section between the ICD-10 and ICD-11, mapping was performed for only the topography codes. To ensure the accuracy of the results, manual recoding was independently implemented by 2 authors who both had more than 10 years of coding experience and had received ICD-11 training. Inconsistent results were resolved by consulting a senior coder from the Collaborating Center for the WHO Family of International Classifications in China.

Statistical Standards
Leaf Codes
The ICD-10, ICD-10-CCM, and ICD-11 MMS codes that can be used at the lowest level are called leaf codes. All statistical analyses were based on the leaf codes.

Coding Capacity
This capacity involves the number of leaf codes that can be used in actual coding.

Grouping Changes
Grouping was based on the block structure of the ICD. The equivalent groups included ICD-10 group 1 (malignant neoplasms) and ICD-11 group 3 (malignant neoplasms, except...
for lymphoid, hematopoietic, central nervous system, or related tissues), ICD-10 group 2 (in situ neoplasms) and ICD-11 group 4 (in situ neoplasms, except for lymphoid, hematopoietic, central nervous system, or related tissues), and ICD-10 group 3 (benign neoplasms) and ICD-11 group 5 (benign neoplasms, except for lymphoid, hematopoietic, central nervous system, or related tissues). If an ICD-10/ICD-10-CCM leaf code was not classified into the equivalent ICD-11 group, it was considered to have undergone a grouping change.

**Expression Ability**

For each ICD-10-CCM code, we identified the best-matching ICD-11 MMS leaf code. When all the clinical details in the diagnosis were expressed without redundant information, the code was considered fully represented. This study also defined the synonyms contained in the Foundation Component as a full representation.

**Statistical Analysis**

All data were analyzed with SPSS (version 25.0, IBM). The changes in coding capacity, groups, and expression ability between the ICD-10 and ICD-11 were described as rates and percentages. The chi-square test was used to determine the difference in grouping changes when mapped to the ICD-11 between the ICD-10 and ICD-10-CCM. The chi-square test was also used to analyze the difference in full expression ability among the 4 ICD-10-CCM groups, and the Bonferroni method was used for pairwise comparisons. The chi-square test was used to analyze the difference in full expression ability between the changed and unchanged groups mapped to the ICD-11. The $\phi$ correlation coefficient was used to analyze the correlation between expression ability and grouping changes. Statistical significance was set at $P<.05$.

**Table 1.** Examples of the International Classification of Diseases, Tenth Revision (ICD-10), Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (ICD-10-CCM), and International Classification of Diseases, Eleventh Revision (ICD-11) complete neoplasm codes.

<table>
<thead>
<tr>
<th>Neoplasm</th>
<th>ICD-10</th>
<th>ICD-10-CCM</th>
<th>ICD-11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topography code</td>
<td>Morphology code</td>
<td>Topography code</td>
<td>Morphology code</td>
</tr>
<tr>
<td>Adenocarcinoma of the common bile duct, stage III, diagnosis confirmed by histology</td>
<td>C24.0</td>
<td>8140/3</td>
<td>C24.003</td>
</tr>
</tbody>
</table>

In the ICD-10, C24.0 represents a malignant neoplasm of the extrahepatic bile duct, and 8140/3 represents adenocarcinoma. In the ICD-10-CCM, C24.003 represents a malignant neoplasm of the common bile duct, and 81400/3 represents adenocarcinoma. In the ICD-11, 2C15.0 represents adenocarcinoma of the distal bile duct, XS6H represents stage III, and XY9Q represents a diagnosis confirmed by histology. We used the the ICD-11 coding tool website [10].

**Coding Capacity**

Categories, subcategories, and leaf codes can be used for statistics, but only leaf codes can be used for actual coding. In the ICD-10, chapter 2 includes 759 leaf codes [7]. In the ICD-10-CCM, the topography codes are basically expanded by refining the sites, with 3634 leaf codes for neoplasms. Taking malignant neoplasm of the nasal cavity (C30.0) as an example, the ICD-10-CCM contains 5 additional leaf codes, such as a malignant neoplasm of nasal cartilages (C30.001) and a malignant neoplasm of the nasal concha (C30.002).

In the ICD-11, chapter 2 includes 1037 leaf codes [8]. Compared with the ICD-10, the number of ICD-11 leaf codes was expanded by 36.6%, which is conducive to better granularity of statistics and classification.

**Classification Feature**

The ICD-11 has readjusted the neoplasm groups. In the ICD-10, chapter 2 was divided into 4 groups: malignant neoplasms, in situ neoplasms, benign neoplasms, and neoplasms of uncertain or unknown behavior. In the ICD-11, chapter 2 was increased to 7 groups (Table 2). The newly added ICD-11 group 1 includes all neoplasms of the brain and central nervous system, regardless of behavior, and the newly added ICD-11 group 2 includes all hematopoietic and lymphoid tissues. The ICD-10 group 4 was split into 2 separate groups: ICD-11 group 6 (neoplasms of
uncertain behavior) and ICD-11 group 7 (neoplasms of unknown behavior).

**Table 2.** Grouping changes between the International Classification of Diseases, Tenth Revision (ICD-10) and the International Classification of Diseases, Eleventh Revision (ICD-11).

<table>
<thead>
<tr>
<th>Classification system</th>
<th>ICD-11</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
<th>Group 7</th>
<th>Other chapters</th>
<th>No mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ICD-10 codes, n</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td>23(^a)</td>
<td>91(^a)</td>
<td>352</td>
<td>N/A(^b)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>8(^a)</td>
<td>2</td>
</tr>
<tr>
<td>Group 2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>57</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Group 3</td>
<td>11(^a)</td>
<td>N/A</td>
<td>1(^a)</td>
<td>N/A</td>
<td>146</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1(^a)</td>
<td>N/A</td>
</tr>
<tr>
<td>Group 4</td>
<td>11(^a)</td>
<td>18(^a)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>48(^a)</td>
<td>51(^a)</td>
<td>2(^a)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>ICD-10-CCM codes, n</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td>93(^a)</td>
<td>499(^a)</td>
<td>1252</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>17(^a)</td>
<td>3</td>
</tr>
<tr>
<td>Group 2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>169</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Group 3</td>
<td>118(^a)</td>
<td>N/A</td>
<td>1(^a)</td>
<td>N/A</td>
<td>796</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>3(^a)</td>
<td>N/A</td>
</tr>
<tr>
<td>Group 4</td>
<td>85(^a)</td>
<td>60(^a)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>316(^a)</td>
<td>149(^a)</td>
<td>18(^a)</td>
<td>55</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Indicates grouping changes. The ICD-10 and ICD-10-CCM no mapping codes were not included in the statistical analysis of this study. Hence, 755 ICD-10 codes and 3576 ICD-10-CCM codes were used for percentage and chi-square analyses.

\(^b\)N/A: not applicable.

In total, 27.4% (207/755) of ICD-10 codes underwent grouping changes. Among them, 150 codes were migrated by 1 group, 56 by 2 groups, and one by 3 groups. In the ICD-10-CCM, 38% (1359/3576) leaf codes underwent grouping changes (Table 2). A chi-square test revealed significant differences in grouping changes between the ICD-10 and ICD-10-CCM ($\chi^2_{1}=30.3$; $P<.001$).

**Expression Features**

**Expression Ability**

Because of the classification changes, 58 of the 3634 ICD-10-CCM codes for neoplasms could not be recoded. In total, 3576 codes were included in the manual recoding study. The results of 2 separate recodings showed that 6% (213/3576) codes were mapped inconsistently. A total of 32 stem codes were inconsistent, and 181 extension codes differed. The 213 codes were all identified after consultation with the senior coder from the Collaborating Center for the WHO Family of International Classifications in China.

The final results showed that 16.6% (594/3576) codes were fully represented without postcoordination, 51.2% (1830/3576) codes were fully characterized with postcoordination, and the remaining 32.2% (1152/3576) codes were only partially described (Table 3). In addition, 7% (252/3576) codes were fully represented when using uniform resource identifiers (URIs).

**Table 3.** Comparison of the International Classification of Diseases, Eleventh Revision (ICD-11) expression ability among the 4 Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (ICD-10-CCM) groups.

<table>
<thead>
<tr>
<th>ICD-10-CCM group</th>
<th>Total codes, N</th>
<th>Full representation</th>
<th>Partial representation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>With PC(^a) (a), n</td>
<td>Without PC (b), n</td>
</tr>
<tr>
<td>Group 1</td>
<td>1861</td>
<td>935</td>
<td>406</td>
</tr>
<tr>
<td>Group 2</td>
<td>169</td>
<td>106</td>
<td>44</td>
</tr>
<tr>
<td>Group 3</td>
<td>918</td>
<td>501</td>
<td>73</td>
</tr>
<tr>
<td>Group 4</td>
<td>628</td>
<td>288</td>
<td>71</td>
</tr>
<tr>
<td>Total</td>
<td>3576</td>
<td>1830</td>
<td>594</td>
</tr>
</tbody>
</table>

\(^a\)PC: postcoordination.

Based on the 4 ICD-10-CCM groups, ICD-11 significantly differed in expression ability ($\chi^2_{3}=93.7$; $P<.001$). The Bonferroni method showed that the ICD-11 had the most robust expression ability in the ICD-10-CCM group 2, followed by the ICD-10-CCM group 1, and there was no significant difference between the remaining 2 groups.

Based on whether grouping changes occurred during mapping, 73.1% (1620/2217) ICD-10-CCM codes in the unchanged group...
and 59.2% (804/1359) ICD-10-CCM codes in the changed group were fully expressed, and the difference was significant ($\chi^2 = 74.7; P < .001$). Expression ability had a negative correlation with grouping changes ($r = -.144; P < .001$).

**Expression Flexibility**

The expression flexibility of the ICD-11 is reflected in many aspects. For instance, in the ICD-10, the subcategory “.8” generally describes overlapping neoplasm sites, while in the ICD-11, several methods are used. Specifically, the ICD-11 uses multiple extension codes. Sometimes, only 1 extension code is used, such as XA4YW8 (overlapping sites of the esophagus). Occasionally, the ICD-11 describes this condition through stem codes, such as 2B71.0 (adenocarcinoma of the esophago gastric junction). ICD-11 classifications can also be represented through URIs, for instance, http://id.who.int/icd/entity/419755630 (Kaposi sarcoma of multiple organs).

Expression flexibility is also reflected in the additional option of postcoordination, which can meet the different requirements of most hospitals for clinical phenotype mining. For example, a patient with hepatocellular carcinoma in the left lobe of the liver that causes chronic intermittent cancer pain and tumor anemia would be coded as 2C12.02&XA5766/MG30.10&XT5G/3A71.0. Postcoordination fully expresses the clinical phenotypes and demonstrates the relationship between diseases and clinical phenotypes.

**Expression of Postcoordination**

According to the ICD-10-CCM/ICD-11 mapping table, 60.5% (2164/3576) of codes had postcoordination, and the average amount of postcoordination per code was approximately 0.7. Regarding the type of postcoordination, the proportion of specific anatomy (1907/3576, 53.3%) was the highest, followed by histopathology (201/3576, 5.6%), and the remaining dimensions were basically less than 2% (Table 4).

Table 4. Expression of postcoordination in the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision/International Classification of Diseases, Eleventh Revision (ICD-10-CCM/ICD-11) mapping table.

<table>
<thead>
<tr>
<th>Number and dimensions of postcoordination</th>
<th>Codes, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1412 (39.5)</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Specific anatomy</td>
<td>1750 (48.9)</td>
</tr>
<tr>
<td>Histopathology</td>
<td>176 (4.9)</td>
</tr>
<tr>
<td>Alternative severity 2</td>
<td>60 (1.7)</td>
</tr>
<tr>
<td>Laterality</td>
<td>5 (0.1)</td>
</tr>
<tr>
<td>Course of the condition</td>
<td>3 (0.1)</td>
</tr>
<tr>
<td>Stem codes</td>
<td>3 (0.1)</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Specific anatomy/specific anatomy</td>
<td>118 (3.3)</td>
</tr>
<tr>
<td>Specific anatomy/histopathology</td>
<td>17 (0.5)</td>
</tr>
<tr>
<td>Specific anatomy/laterality</td>
<td>8 (0.2)</td>
</tr>
<tr>
<td>Histopathology/alternative severity 2</td>
<td>8 (0.2)</td>
</tr>
<tr>
<td>Alternative severity 2/stem codes</td>
<td>2 (0.1)</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Specific anatomy/specific anatomy/specific anatomy</td>
<td>11 (0.3)</td>
</tr>
<tr>
<td>Specific anatomy/specific anatomy/laterality</td>
<td>3 (0.1)</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

The different coding structure of the ICD-11 provides advantages in terms of coding granularity, coding capacity, and expression flexibility. According to the mapping tables, the grouping changes between the ICD-10 and ICD-10-CCM differed ($\chi^2 = 30.3; P < .001$). Meanwhile, neither the ICD-10 group 2 nor the ICD-10-CCM group 2 exhibited group migration when mapped to the ICD-11. In this study, ICD-11 expression ability (67.8%) was slightly higher than that in other studies (60%) [12, 13]. Among the 4 ICD-10-CCM groups, there were significant differences in expression ability ($\chi^2 = 93.7; P < .001$), with ICD-10-CCM group 2 having the highest expression ability. The expression ability negatively correlated with grouping changes ($r = -.144; P < .001$).

The expression ability of the ICD-11 was still underestimated. There were many reasons why 1152 ICD-10-CCM codes did not have full representation, some of which can be avoided in actual cases. First, of the 296 ICD-10-CCM codes mapped to the ICD-11 group 1, only 26 (8.8%) codes were fully expressed.
which is far below the overall level. Among them, 217 codes cannot be fully expressed because most of the stem codes in ICD-11 group 1 do not include behavior. However, in coding actual cases with morphological types, some stem codes, including behavior, would be used instead of the residual category. Therefore, in actual coding, the expression ability of this group would be higher than that determined in this study.

Second, 714 codes could not be fully expressed due to site classification. Fortunately, 109 codes detailing specific sites may be resolved by refining the value set of extension codes. Examples include the frenulum of the upper lip, the ileocecal valve, and the rectouterine recess. However, there were also some codes for which the classification was different, such as peripheral nervous system neoplasms. This condition could be solved by coding the actual cases. In addition, some words, such as canceration of the gastric stump and cervical stump, need to be addressed.

Third, URIs can supplement the function of the ICD-11 MMS codes, especially for some diseases that cannot be identified by the ICD-11 codes. For instance, URIs (http://id.who.int/icd/entity/1595913346) make classic Kaposi sarcoma classifiable. The new coding system can also make other diseases identifiable [14-16], such as chronic pain and rare diseases. In summary, the ICD-11 has advantages in terms of actual coding and can address the expression needs of neoplasms.

The ICD-11 Vs the ICD-10

Compared to the ICD-10, the ICD-11 has undergone significant changes from design to use. First, the design purpose of the ICD-11 has changed. For more than a century, the ICD has been the basis for comparable statistics on causes of mortality and morbidity between places and over time. As a statistically friendly classification system, when the ICD-10 is used for clinical term records and diagnosis-related grouping, it cannot meet practical needs, resulting in various expanded versions of the WHO ICD-10 in multiple countries, which often leads to inconsistent statistical standards for data. In contrast, the ICD-11 is a clinically friendly classification system that meets diverse goals beyond mere health statistics, including clinical term records, patient safety and quality, reimbursement, decision support, and more. The ICD-11 is entirely digital, terminology is coded with the coding tool and application programming interface, and it has a semantic knowledge base; these features are beneficial for standardized data collection. In short, the ICD-11 has benefits in terms of obtaining statistics as well as multiaxial coding, coding granularity [4,17,18] and standardization, achieving the integration of terminology and classification. As shown in Table 1, although the ICD-10-CCM codes are expanded and refined at the leaf code level, the capture of clinical details is still weaker compared to the ICD-11 codes.

Second, the design concept of the ICD-11 is different. The ICD-10 is an independent classification system, and the standard terminology set for the ICD-10 has been developed independently. Differently, the ICD-11 has good semantic interoperability with other classification systems through harmonized methods. The Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) is considered one of the most comprehensive clinical terminologies in the world [19]. The ICD-11 Foundation Component, which includes semantic network concepts and their relationships, is organized around the Common Ontology from a subset of the SNOMED CT [20,21]. The Common Ontology has been harmonized with ICD text definitions, primarily from the SNOMED CT clinical findings hierarchy (findings, disorders, and disease) and secondarily from other hierarchies (situations, events, social context, and so on) [22]. The rich Foundation Component has approximately 80,000 entries and 40,000 synonyms [5]. The ICD-11 linearizations, including the ICD-11 MMS and the International Classification of Diseases for Oncology (ICD-O), are subsets derived from the Foundation Component. The ICD-11 integrates the morphology section of the ICD-O, ICD-O linearization, and tumor node metastasis classification, and the histopathology codes of the ICD-11 are also compatible with the ICD-O [4]. Additionally, the ICD-11 integrates numerous clinical terminologies from some expanded versions of the WHO ICD-10. Compared to other classifications, the ICD-11 has advantages in terms of concept coverage and compatibility.

Third, the ICD-11 has stronger logical links between codes. If multiple codes are needed for disease expression, there is a lack of practical connections between ICD-10 codes. In the ICD-11, precoordination contains the site and morphology, and the postcoordination of morphology and the clinical phenotype is linked to stem codes through an ampersand (&) and slash (/).

In addition, the ICD-11 contains rich dimensions of postcoordination. Compared to the ICD-10, which provides information only about topography and morphology, the ICD-11 can also include other dimensions of postcoordination, such as stage, grading, laterality, and the diagnostic method. The ICD-11 coding system is more conducive to unifying national cancer registration in these dimensions. However, due to the limitations of the ICD-10 coding structure, these dimensions of postcoordination were missing when mapped to the ICD-11, with only 0.7 postcoordination per code, and its expression ability cannot be truly reflected.

Moreover, the ICD-11 has more expression flexibility. The ICD-10 coding system is fixed and single. The ICD-11 provides 2 sets of codes: ICD-11 MMS codes and URIs. In ICD-11 MMS codes, a stem code can be used alone or with optional extension codes. A URI is a string of characters that uniquely identifies a particular entity. Therefore, a coding system with good expression flexibility can meet the coding granularity and clinical phenotype mining requirements of hospitals at different levels.

Finally, the ICD-11 has different update mechanisms. The ICD-10 updates relatively slowly. In 1999, the WHO established the Update Advisory Committee as the only authoritative body revising the ICD-10. In 2005, the WHO officially published the second edition of the ICD-10 and continued to issue revisions to the relevant content of the ICD-10 through official channels every year. The ICD-11 provides a web-based coding tool instead of paper environments, which is highly beneficial for timely updates. It has a real-time updated orange version and an annually updated blue version.
Overall, the eleventh revision is more extensive than any other revision since the sixth in 1948. These design-level features make the ICD-11 competitive with other classification systems.

**Challenges of Replacement**

The improvements described above also pose challenges for replacing the ICD-10 with the ICD-11. Compared to the ICD-10, the ICD-11 leaf codes for neoplasms have expanded by 36.6%, and there are also approximately 16,000 extension codes. Hence, the expression ability of the ICD-11 for clinical details far exceeds that of the ICD-10 and ICD-10-CCM. However, codes with better granularity require clinicians to record diagnoses and treatments in greater detail. It is essential to carry out the necessary interventions to enhance medical record documentation according to ICD-11 before or simultaneously with country-wide implementation [23]. Clinicians will be required to have a good understanding of the ICD-11. Otherwise, the best classification system is just a decoration.

Second, coders are accustomed to using the old classification system and need time to familiarize themselves with the new coding system and tool. Studies have shown that some coders have difficulties coding on the web due to a lack of familiarity with software [24,25]. In this study, 2 skilled coders who received ICD-11 training still had 213 inconsistent codes. Web-based coding requires high professional ability. Studies have also shown that the reliability and accuracy of ICD-11 coding are lower than those of ICD-10 coding [23-25]. In the early stages of ICD-11 promotion, countries involved in the transition need to conduct ICD-11 training and transform coders’ coding concepts. To maintain coding accuracy, developing high-quality ICD-11 training materials, training processes, and local ICD-11 guidelines is effective [23,26].

Moreover, many countries have applied the ICD-10 in different fields, such as mortality and morbidity statistics, diagnosis-related grouping, and cancer registries. There is also a potential challenge with IT systems being able to adopt a new classification that allows for unlimited width in fields. Stakeholders must upgrade their information systems to meet the needs of the ICD-11, which requires a significant amount of time and money to hire medical, IT, and management personnel to complete this transformation. The experience of the United States in replacing ICD versions can provide a better reference. In the United States, because of significant opposition and reservations expressed by stakeholders, it took 6 years from the adoption of the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) to complete the transition [27], and the costs associated with the transition were estimated to be between US $475 million and US $1.5 billion dollars, including training, productivity losses, and system changes [28].

Importantly, the ICD-10 and ICD-11 will coexist for a long time before replacement, which contributes to solving some problems. The parallel implementation of the 2 classification systems can ensure government policy continuity. In January 2022, the pilot application program to promote ICD-11 was launched in China, with 59 large general public hospitals from all provinces participating. However, ICD-11 is not fully understood by other hospitals. The impacts of the new classification system implementation included coding accuracy, ICD version mapping, and more [29]. The ICD version mapping can be used for interoperability between coded data sets [30]. Hospitals do not require dual coding and use ICD-11 to report data. If a hospital needs dual coding, a mapping table can significantly reduce the time, at least by half, spent on duplicate coding. A map can also maintain the same accuracy of ICD-11 coding for all hospitals and alert coders if there are any coding errors. Usually, a high-quality map requires substantial manual curation, and some studies use algorithmic mapping approaches, such as sequential mapping [30]. Some scholars have developed hybrid methods, such as automatic mapping and manual review [31], semiautomatic mapping, and manual evaluation [32]. Although automatic coding has many benefits, supervised and manual mapping are still necessary [33]. Undeniably, there may be some issues with the use of mapping tables. A study has shown that 14.5% of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes used by internists, when mapped to the ICD-10-CM, resulted in potential clinical inaccuracies [34]. In general, a high-quality map would contribute to the smooth transition of ICD-11, and multiple mapping methods can be explored. This study confirms the feasibility of using manual mapping tables for neoplasms between the ICD-10-CCM and ICD-11. To date, the results of some studies on the transition to the ICD-11 have also been optimistic [17,18].

In addition, although the WHO provides an ICD-11 coding tool, countries must continue optimizing this tool. A mapping table can improve the intelligence of the coding tool. If the terms are the same as those in the mapping table, the coding tool can automatically load the coding cluster without selecting postcoordination one by one. The accuracy and efficiency of ICD-11 coding largely depend on how well the coding tool is optimized. It is necessary to continuously optimize the coding tool before fully promoting the ICD-11.

**Limitations**

This study had some limitations that must be considered. First, this study focused only on the clinical condition of neoplasms. However, the chapter on neoplasms is quite complex and one of the most varied chapters on other conditions, with significant changes. Moreover, patients with cancer may experience different manifestations, including chronic, surgical, and emergency conditions. For example, a patient with long-term chronic hepatitis B causing cirrhosis and liver cancer was admitted to the hospital for emergency surgery due to a tumor rupture. Furthermore, the burden of cancer in China continues to grow with the aging population. According to statistics, the crude cancer mortality rate increased from 108.3 per 100,000 individuals in the 1990-1992 period to 170.1 per 100,000 individuals in 2015 [35]. Significantly, the research method is well thought out. Manual mapping can provide a good understanding of the new ICD-11 features and help individuals familiarize themselves with the characteristics of the new classification systems. As a method, manual mapping can be extended to other conditions. More conditions can be gradually incorporated, especially emergency and surgical conditions, in the future.
Second, this study used only manual mapping, which is time-consuming and laborious. Other methods combined with manual mapping can be continuously explored, especially automatic mapping between the ICD-10-CCM and ICD-11, including algorithmic mapping and machine learning approaches.

Conclusion

Neoplasm classification has undergone many improvements in the ICD-11, especially the new coding system, improved expression ability, and good semantic interoperability. The new coding system provides advantages in coding granularity, coding capacity, and expression flexibility. Moreover, 67.8% of ICD-10-CCM neoplasm codes can be fully represented by the ICD-11, and expression ability negatively correlates with grouping changes ($r=-0.144$, $P<.001$). The more significant the changes in a new classification system are, the less information can be expressed when mapped to other databases. The use of URIs and maintenance mechanisms can increase the expression ability for coding in actual cases. In addition, the good semantic interoperability of the ICD-11, integrating numerous clinical terminologies from the SNOMED CT, the ICD-O, the tumor node metastasis classification, and expanded versions of the WHO ICD-10, makes the ICD-11 competitive with other classification systems. The transition to the ICD-11 will inevitably bring numerous challenges for clinicians, coders, policy makers and IT technicians, and many preparations will be necessary.

Acknowledgments

The authors would like to thank Dr Xiaoding Liu from the Department of Pathology at Peking Union Medical College Hospital for providing information about pathology and answering our questions. This work was supported by the Fundamental Research Funds for the Central Universities (grant 2022FZZX06-09), and the Zhejiang Kangenbei Hospital-managed soft science research project (grant 2023ZHA-KEB305).

Conflicts of Interest

None declared.

References


Abbreviations

ICD-10: International Classification of Diseases, Tenth Revision
ICD-10-CCM: Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
ICD-11 MMS: International Classification of Diseases, Eleventh Revision, Mortality and Morbidity Statistics
ICD-11: International Classification of Diseases, Eleventh Revision
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-O: International Classification of Diseases for Oncology
SNOMED CT: Systematized Nomenclature of Medicine-Clinical Terms
URI: uniform resource identifier
WHO: World Health Organization
Application of AI in Sepsis: Citation Network Analysis and Evidence Synthesis

MeiJung Wu¹,², MSc; Md Mohaimenul Islam³, PhD; Tahmina Nasrin Poly¹, MSc; Ming-Chin Lin¹,⁴,⁵, PhD

¹Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
²Department of Nursing, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
³Department of Outcomes and Translational Sciences, College of Pharmacy, The Ohio State University, Columbus, OH, United States
⁴Taipei Neuroscience Institute, Taipei Medical University, Taipei, Taiwan
⁵Department of Neurosurgery, Shuang Ho Hospital, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Ming-Chin Lin, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
250 Wuxing St
Xinyi District
Taipei, 110
Taiwan
Phone: 886 66202589
Email: arbiter@tmu.edu.tw

Abstract

Background: Artificial intelligence (AI) has garnered considerable attention in the context of sepsis research, particularly in personalized diagnosis and treatment. Conducting a bibliometric analysis of existing publications can offer a broad overview of the field and identify current research trends and future research directions.

Objective: The objective of this study is to leverage bibliometric data to provide a comprehensive overview of the application of AI in sepsis.

Methods: We conducted a search in the Web of Science Core Collection database to identify relevant articles published in English until August 31, 2023. A predefined search strategy was used, evaluating titles, abstracts, and full texts as needed. We used the Bibliometrix and VOSviewer tools to visualize networks showcasing the co-occurrence of authors, research institutions, countries, citations, and keywords.

Results: A total of 259 relevant articles published between 2014 and 2023 (until August) were identified. Over the past decade, the annual publication count has consistently risen. Leading journals in this domain include Critical Care Medicine (17/259, 6.6%), Frontiers in Medicine (17/259, 6.6%), and Scientific Reports (11/259, 4.2%). The United States (103/259, 39.8%), China (83/259, 32%), United Kingdom (14/259, 5.4%), and Taiwan (12/259, 4.6%) emerged as the most prolific countries in terms of publications. Notable institutions in this field include the University of California System, Emory University, and Harvard University. The key researchers working in this area include Ritankar Das, Chris Barton, and Rishikesan Kamaleswaran. Although the initial period witnessed a relatively low number of articles focused on AI applications for sepsis, there has been a significant surge in research within this area in recent years (2014–2023).

Conclusions: This comprehensive analysis provides valuable insights into AI-related research conducted in the field of sepsis, aiding health care policy makers and researchers in understanding the potential of AI and formulating effective research plans. Such analysis serves as a valuable resource for determining the advantages, sustainability, scope, and potential impact of AI models in sepsis.


KEYWORDS
AI; artificial intelligence; bibliometric analysis; bibliometric; citation; deep learning; machine learning; network analysis; publication; sepsis; trend; visualization; VOSviewer; Web of Science; WoS
Introduction

Sepsis is a life-threatening medical emergency [1] affecting approximately 48.9 million individuals globally each year and potentially contributing to over 11 million deaths [2]. Previous studies indicated that sepsis-related hospitalization can result in fatal outcomes in 30%-50% of cases [3,4]. However, prompt stratification and the timely administration of specific treatments have the potential to lower sepsis-related mortality. Identifying sepsis at an early stage can be challenging due to the complex pattern of the disease [5,6] and the diversity of the septic population [7].

Artificial intelligence (AI) has piqued interest in its excellent potential to stratify patients with a high risk of sepsis [8]. In recent times, AI models have seen widespread application in the prediction of sepsis and have shown superior performance compared with conventional statistical methods [9,10]. Yet, no study has shed light on the variety of AI applications and their potential and limitations in sepsis through a scientific consolidation of knowledge. Bibliometric analysis aids researchers in comprehending specific research fields, a crucial aspect for guiding both future research endeavors (eg, what else should we know) and practical implementation (eg, what should we do) [11]. This research aims to address the following questions, with the intent of advancing the previous research on the application of AI in sepsis: (1) What countries, institutions, sources, and documents have demonstrated the highest productivity within the realm of AI applied to sepsis? (2) What are the hot research topics and themes of research in the application of AI in sepsis? (3) What methods are mainly applied in the existing body of literature? (4) What types of limitations appeared in the existing literature regarding the application of AI in sepsis? and (5) What are the literature gaps and future research agendas?

In this study, we could systematically investigate shifts in publication growth, offering more valuable insights to fellow researchers and policy makers engaged in priority setting and assessment.

Methods

Data Source

We leveraged extracted data from the Web of Science Core Collection as of August 31, 2023. We used Web of Science for its comprehensive coverage across multiple databases, comprising a wide range of bibliographic indicators and literature from various disciplines. Using a predefined search strategy, we intended to include all relevant literature for bibliometric analysis. We used the following key words: artificial intelligence OR computational intelligence OR deep learning OR computer aided OR machine learning OR support vector machine OR data learning OR artificial neural network OR digital image OR convolutional neural network OR evolutionary algorithms OR feature learning OR reinforcement learning OR big data OR image segmentation OR hybrid intelligent system OR hybrid intelligent system OR recurrent neural network OR natural language processing OR Bayesian network OR Bayesian learning OR random forest OR evolutionary algorithms OR multiagent system AND sepsis. The collected records contained essential attributes, including publication date, authorship, institutional affiliation, geographic origin, and cited references. This data set served as the foundation for our subsequent analytical investigations.

Inclusion and Exclusion Criteria

The titles and abstracts underwent initial screening by 2 independent authors (MW and TNP). If there was uncertainty from one reviewer regarding whether the article met the inclusion criteria, it was included for a thorough full-text review. Following this, both authors independently assessed the full text, and any differences in opinion were resolved through consensus with the research team. We considered studies for inclusion if they met the following criteria: (1) they were written in English, and (2) they applied AI models in the context of sepsis. In this study’s screening process, we included research or review articles published in peer-reviewed journals, conference proceedings, reviews, and early access articles. We excluded studies if they were published as letters, editorials, book chapters, or books.

Data Collection and Preprocessing

To ensure compatibility with Bibliometrix and VOSviewer [12], we saved the data in the “**.txt” format, a format recognized by both tools for conducting analyses. Our data set encompasses a comprehensive range of information, including titles, list of authors, name of countries, list of institutions, abstracts, keywords, name of journals, and publication dates.

Statistical Analysis

Bibliometrix and VOSviewer tools were used to uncover the knowledge structure, most influential countries, research hot spots, and productive authors, along with various bibliometric insights. The processed data were uploaded into these bibliometric tools, and analysis was conducted based on the information included within the data documents [13]. Afterward, we generated network maps among journals, authors, countries, and institutions, where individual points symbolized authors, countries, or institutions. Moreover, connected lines in the network maps depicted the relationships between these entities. Larger points and more robust lines indicated a higher number of articles and more substantial collaborative relationships, respectively [14,15].

We computed the annual growth rate of publications. The annual publication count, annual growth, and average growth rate of publications were determined through the following methods:

\[
N_T = \sum_{k=1}^{T} N_k
\]

Where \(N\) is the total number of articles in the current year, and \(N_{k-1}\) is the total number of articles in the previous year.

Furthermore, we conducted an analysis of publication trends based on the following criteria: the top 10 most prolific countries, institutions, journals, authors, and studies in this area. The rankings of countries, institutions, journals, and authors were determined based on the number of articles.
Results

Distribution of Articles by Publication Year
The initial search yielded 327 articles focused on the application of AI in sepsis. After applying predefined inclusion criteria, 68 articles were deleted, leaving 259 articles for the final analysis (Figure 1).

Over time, there has been a substantial rise in the number of publications in this field. Notably, the yearly publication number increased from just 2 articles in 2014 to 72 articles in 2022. It is important to note that before 2018, the yearly publication count did not cross 10 articles. The calculated annual growth rate was found to be 44.81% (Figure 2).

Figure 1. A diagram for the detailed selection criteria and bibliometric analysis steps of applying artificial intelligence to sepsis in the Web of Science Core Collection database.

![Diagram of selection criteria](image)

Figure 2. Trends in the number of publications on the application of artificial intelligence to the study of sepsis from 2014 to 2023 (August).

![Trend graph](image)

Distribution of Source Journals
A total of 122 journals published articles on the application of AI in sepsis. Among them, the *Critical Care Medicine* journal was the most productive, having published 6.6% (17/122) of articles in this domain (Table 1). *Frontiers in Medicine*, *Scientific Reports*, and the *American Journal of Respiratory and Critical Care Medicine* were in the second, third, and fourth positions, publishing 17, 11, and 7 articles, respectively, on this topic. However, the top 10 journals published 86 articles, accounting for 33.2% (86/259) of all publications in this area.
Table 1. The top 10 journals with publications on the application of artificial intelligence in sepsis from 2014 to August 2023.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Journal</th>
<th>Country</th>
<th>Category</th>
<th>Publication frequency, n (%)</th>
<th>Impact factor in 2022</th>
<th>5-year impact factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Critical Care Medicine</td>
<td>United States</td>
<td>Engineering, electrical, and electronics</td>
<td>17 (6.6)</td>
<td>8.8</td>
<td>8.4</td>
</tr>
<tr>
<td>2</td>
<td>Frontiers in Medicine</td>
<td>Switzerland</td>
<td>Multidisciplinary science</td>
<td>17 (6.5)</td>
<td>3.9</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>Scientific Reports</td>
<td>United Kingdom</td>
<td>Multidisciplinary science</td>
<td>11 (4.2)</td>
<td>4.6</td>
<td>4.9</td>
</tr>
<tr>
<td>4</td>
<td>American Journal of Respiratory and Critical Care Medicine</td>
<td>United States</td>
<td>Multidisciplinary science</td>
<td>7 (2.7)</td>
<td>24.7</td>
<td>21.9</td>
</tr>
<tr>
<td>5</td>
<td>Frontiers in Immunology</td>
<td>Switzerland</td>
<td>Clinical neurology</td>
<td>6 (2.3)</td>
<td>7.3</td>
<td>8.0</td>
</tr>
<tr>
<td>6</td>
<td>Intensive Care Medicine</td>
<td>United States</td>
<td>Neurosciences</td>
<td>6 (2.3)</td>
<td>38.9</td>
<td>27</td>
</tr>
<tr>
<td>7</td>
<td>Journal of the American Medical Informatics Association</td>
<td>United States</td>
<td>Computer science and interdisciplinary applications</td>
<td>6 (2.3)</td>
<td>6.4</td>
<td>6.3</td>
</tr>
<tr>
<td>8</td>
<td>PLoS One</td>
<td>United States</td>
<td>Neurosciences</td>
<td>6 (2.3)</td>
<td>3.7</td>
<td>3.8</td>
</tr>
<tr>
<td>9</td>
<td>BMC Medical Informatics and Decision Making</td>
<td>United Kingdom</td>
<td>Engineering and multidisciplinary</td>
<td>5 (1.9)</td>
<td>3.5</td>
<td>3.9</td>
</tr>
<tr>
<td>10</td>
<td>Computers in Biology and Medicine</td>
<td>United States</td>
<td>Engineering and biomedical</td>
<td>5 (1.9)</td>
<td>7.7</td>
<td>6.9</td>
</tr>
</tbody>
</table>

Distribution of Countries and Regions
This study revealed that researchers from 73 countries and regions engaged in research on these subjects and published their work in various international peer-reviewed journals. Out of the total 259 articles, the United States made the most substantial contribution with 103 publications (39.8%), followed by China with 83 publications (32%), United Kingdom with 14 publications (5.4%), and Taiwan with 12 publications (4.6%) (Table 2).

Table 2. The top 10 countries and regions with publications on the application of artificial intelligence in sepsis from 2014 to August 2023.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Articles, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>United States</td>
<td>103 (39.8)</td>
</tr>
<tr>
<td>2</td>
<td>China</td>
<td>83 (32)</td>
</tr>
<tr>
<td>3</td>
<td>United Kingdom</td>
<td>14 (5.4)</td>
</tr>
<tr>
<td>4</td>
<td>Taiwan</td>
<td>12 (4.6)</td>
</tr>
<tr>
<td>5</td>
<td>India</td>
<td>11 (4.2)</td>
</tr>
<tr>
<td>6</td>
<td>Netherlands</td>
<td>10 (3.9)</td>
</tr>
<tr>
<td>7</td>
<td>Australia</td>
<td>8 (3.1)</td>
</tr>
<tr>
<td>8</td>
<td>Canada</td>
<td>8 (3.1)</td>
</tr>
<tr>
<td>9</td>
<td>Spain</td>
<td>7 (2.7)</td>
</tr>
<tr>
<td>10</td>
<td>Germany</td>
<td>7 (2.7)</td>
</tr>
</tbody>
</table>

Distribution of Institutions
Table 3 shows the top 10 most productive institutes that used AI applications in sepsis. The University of California system (22/259 articles, 8.5%) ranked first among all research institutions, followed by Emory University (10/259 articles, 3.9%), Harvard University (10/259 articles, 3.9%), and Central South University (8/259 articles, 3.1%).

Figure 3 shows the institution cooperation network of 117 institutions that published at least 1 article.
Table 3. The top 10 institutes with publications on the application of artificial intelligence in sepsis from 2014 to August 2023.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Institutions</th>
<th>Country</th>
<th>Publications, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>University of California system</td>
<td>United States</td>
<td>22 (8.5)</td>
</tr>
<tr>
<td>2</td>
<td>Emory University</td>
<td>United States</td>
<td>10 (3.9)</td>
</tr>
<tr>
<td>3</td>
<td>Harvard University</td>
<td>United States</td>
<td>10 (3.9)</td>
</tr>
<tr>
<td>4</td>
<td>Central South University</td>
<td>China</td>
<td>8 (3.1)</td>
</tr>
<tr>
<td>5</td>
<td>Dascena Inc</td>
<td>United States</td>
<td>8 (3.1)</td>
</tr>
<tr>
<td>6</td>
<td>University of Pennsylvania</td>
<td>United States</td>
<td>8 (3.1)</td>
</tr>
<tr>
<td>7</td>
<td>Zhejiang University</td>
<td>China</td>
<td>8 (3.1)</td>
</tr>
<tr>
<td>8</td>
<td>Stanford University</td>
<td>United States</td>
<td>7 (2.7)</td>
</tr>
<tr>
<td>9</td>
<td>Sun Yat-sen University</td>
<td>China</td>
<td>7 (2.7)</td>
</tr>
<tr>
<td>10</td>
<td>Fudan University</td>
<td>China</td>
<td>6 (2.3)</td>
</tr>
</tbody>
</table>

Figure 3. Institution co-operation network.

Distribution of Authors

A total of 259 articles were authored by 1444 individuals, each of whom had at least 1 article to their name. In Table 4, we present the top 10 most prolific authors who conducted and published research in these fields. Ritankar Das secured the top position with 8 articles, closely followed by Chris Barton (6 articles), Rishikesan Kamaleswaran (6 articles), and Suchi Saria (6 articles).

Our analysis shows that 1444 authors have published at least 1 article. The largest set of associated authors consisted of 20 authors in 3 clusters (Figure 4).
Table 4. The top 10 authors with publications on the application of artificial intelligence in sepsis from 2014 to August 2023.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Author</th>
<th>Articles, n</th>
<th>Citations, n</th>
<th>h-index</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Das</td>
<td>8</td>
<td>1417</td>
<td>18</td>
<td>Dascena Inc</td>
</tr>
<tr>
<td>2</td>
<td>Barton</td>
<td>6</td>
<td>2277</td>
<td>25</td>
<td>University of California San Francisco</td>
</tr>
<tr>
<td>3</td>
<td>Kamaleswaran</td>
<td>6</td>
<td>861</td>
<td>14</td>
<td>Emory University</td>
</tr>
<tr>
<td>4</td>
<td>Saria</td>
<td>6</td>
<td>2635</td>
<td>23</td>
<td>Johns Hopkins University</td>
</tr>
<tr>
<td>5</td>
<td>Calvert</td>
<td>5</td>
<td>1152</td>
<td>17</td>
<td>University of California Berkeley</td>
</tr>
<tr>
<td>6</td>
<td>Hoffman</td>
<td>5</td>
<td>957</td>
<td>14</td>
<td>Dascena Inc</td>
</tr>
<tr>
<td>7</td>
<td>Li</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>Sun Yat-sen University</td>
</tr>
<tr>
<td>8</td>
<td>Nemati</td>
<td>5</td>
<td>2402</td>
<td>23</td>
<td>University of California San Diego</td>
</tr>
<tr>
<td>9</td>
<td>Adams</td>
<td>4</td>
<td>118</td>
<td>6</td>
<td>Johns Hopkins University</td>
</tr>
<tr>
<td>10</td>
<td>Davis</td>
<td>4</td>
<td>1253</td>
<td>17</td>
<td>University of Tennessee Health Science Center</td>
</tr>
</tbody>
</table>

Figure 4. The co-authorship network of authors who contributed research on the application of artificial intelligence to sepsis from 2014 to 2023 (August).

Articles Cocitation Analysis

Table 5 shows the top 10 most frequently cited publications. The publication that received the most citations was by Komorowski et al [16], titled “The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care,” published in *Nature Medicine* in 2018 and received a total of 408 citations as of August 31, 2023.
Table 5. Top 10 cited articles in the application of artificial intelligence on sepsis research from 2014 to August 2023.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Author</th>
<th>Journal</th>
<th>Title</th>
<th>Citation, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Komorowski et al [16]</td>
<td>Nature Medicine</td>
<td>The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care</td>
<td>408</td>
</tr>
<tr>
<td>5</td>
<td>Fleuren et al [10]</td>
<td>Intensive Care Medicine</td>
<td>Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy</td>
<td>184</td>
</tr>
<tr>
<td>7</td>
<td>Gultepe et al 2014 [21]</td>
<td>Journal of the American Medical Informatics Association</td>
<td>From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system</td>
<td>101</td>
</tr>
<tr>
<td>8</td>
<td>Giannini et al 2019 [22]</td>
<td>Critical Care Medicine</td>
<td>A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>Mani et al 2014 [24]</td>
<td>Journal of the American Medical Informatics Association</td>
<td>Medical decision support using machine learning for early detection of late-onset neonatal sepsis</td>
<td>97</td>
</tr>
</tbody>
</table>

Co-Occurrence Analysis of Top 100 Keywords

Keywords encapsulate the central themes within a publication and are ideal for examining interconnected areas of research. In this study, we performed co-occurrence analysis to pinpoint the prominent research focal points in the field of AI application in sepsis research, using the top 100 keywords. The extraction and clustering of these top 100 keywords were performed using VOSviewer.

Figure 5 illustrates our use of VOSviewer to create a visual network map, consisting of 6 clusters based on the co-occurrence of the top 100 keywords. The core of this visualization network map is occupied by the following keywords: sepsis (n=138), machine learning (n=122), artificial intelligence (n=35), and deep learning (n=20).
Discussion

Main Findings

This study investigated the citation analyses of AI models in sepsis research, scrutinizing the publication patterns related to their application. This study reports a noteworthy upswing in interest in this subject over the past decade, particularly increasing from 2018 to 2022.

The substantial increase in the use of AI in health care research is noteworthy. For instance, there was an 88.88% increase in the use of AI in health care research after 2012, with only 10 countries contributing to over 96% of the studies [25]. Our findings also indicate an annual growth rate of 44.81% in the use of AI in sepsis-related research. If this growth rate continues in the future, we can anticipate the publication volume in the domain of AI in sepsis-related research doubling approximately every 5 years. This increased number of publications indicates advancements, improved functionality, and the overall progress of AI in sepsis, especially when compared to other application areas.

The number of publications exhibits geographical disparities [26-28]. Recent research indicates a rise in the application of AI in health care research, particularly in developed countries [12,29,30]. These nations are investing increased funds into AI research and the development of AI tools to improve health outcomes. This study aligns with these findings, highlighting an increased number of publications originating from developed countries. However, it underscores the importance of researchers from developing nations stepping forward to contribute toward achieving ultimate health goals. Effective collaboration among clinical experts, AI model developers, and health care providers is essential to addressing the challenges at hand [31].

Selecting the appropriate journals for publication can be a complex decision [32]. Authors take various factors into account when submitting, such as the accessibility of open-access journals and the higher impact factor associated with certain subscription journals [33]. Some authors prefer open-access journals for their widespread availability after publication and the potential for a higher number of citations [34]. Conversely, subscription journals from reputable publishers attract attention due to their high impact factors [35,36]. New researchers often struggle with the dilemma of where to submit their work. This study shows that authors frequently weigh both open-access and higher-impact-factor journals when publishing research related to the application of AI in sepsis.

This study aimed primarily to highlight the most commonly used data sets and algorithms in the current literature. The majority of studies used the Multiparameter Intelligent Monitoring in Intensive Care III (MIMIC-III) data set and supervised machine learning models [37-39]. Notably, Komorowski et al [16] introduced a reinforcement learning model to predict sepsis in patients, demonstrating average reliability levels higher than those of human clinicians. While many studies used extensive training and testing data sets, the
majority focused on single-centered data [39]. To apply these models in real-world clinical settings, external validation becomes necessary.

Strengths and Limitations
This study has several strengths. First, it is the first comprehensive bibliometric analysis that sheds light on the research trends of the application of AI in sepsis, illustrating how this field has evolved. Second, this study gauges productivity in terms of sources, authors, institutions, and countries, while also visualizing word trends. This provides novel and in-depth insights for both researchers and practitioners. This study also has some limitations to address. First, we only collected relevant publications from the Web of Science, a widely used academic resource, for bibliometric analyses [13,40-42]. Nevertheless, using other databases, such as PubMed or Scopus, might have provided slightly varied findings. Second, our inclusion criteria comprised articles published solely in English. However, inclusion of other languages, gray literature, and books might have influenced outcomes, particularly considering diverse cultural perspectives among scholars on the application of AI in sepsis. Finally, relying solely on article titles for the search may pose limitations. However, our aim was to focus on publications specifically addressing the application of AI in sepsis. Therefore, a title screening was deemed more suitable than a broader topic search.

Conclusion
This study aimed to present a comprehensive overview of the use of AI in sepsis through a systematic analysis of existing literature. The findings of this study reveal a noticeable increase in the number of publications over the last 10 years. Until now, developed countries have been the primary contributors in this field. Researchers from developing countries should step forward, leveraging population advantages and core technologies in different regions to foster collaboration.

Leading multidisciplinary science journals, including *Frontiers in Medicine, Scientific Reports*, and the *American Journal of Respiratory and Critical Care Medicine*, emerge as key contributors to this topic based on the volume of published articles. As the application of AI in sepsis research continues to rise, this study serves as a valuable resource for researchers seeking direction and opportunities for collaboration.

Authors' Contributions
All authors contributed to the conception and design of the study; data collection, analysis, and interpretation; drafting and revising the article critically for content; and approval of the final version to be submitted.

Conflicts of Interest
None declared.

References


Abbreviations

AI: artificial intelligence

MIMIC-III: Multiparameter Intelligent Monitoring in Intensive Care III
The advent of digital health technologies has transformed the landscape of health care, influencing the dynamics of the physician-patient relationship. Although these technologies offer potential benefits, they also introduce challenges and complexities that require ethical consideration.

This scoping review aims to investigate the effects of digital health technologies, such as digital messaging, telemedicine, and electronic health records, on the physician-patient relationship. To understand the complex consequences of these tools within health care, it contrasts the findings of studies that use various theoretical frameworks and concepts with studies grounded in relational ethics.

Using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, we conducted a scoping review. Data were retrieved through keyword searches on MEDLINE/PubMed, Embase, IEEE Xplore, and Cochrane. We screened 427 original peer-reviewed research papers published in English-language journals between 2010 and 2021. A total of 73 papers were assessed for eligibility, and 10 of these were included in the review. The data were summarized through a narrative synthesis of the findings.

Digital health technologies enhance communication, improve health care delivery efficiency, and empower patients, leading to shifts in power dynamics in the physician-patient relationship. They also potentially reinforce inequities in health care access due to variations in technology literacy among patients and lead to decreases in patient satisfaction due to the impersonal nature of digital interactions. Studies applying a relational ethics framework have revealed the nuanced impacts of digital health technologies on the physician-patient relationship, highlighting shifts toward more collaborative and reciprocal care. These studies have also explored transitions from traditional hierarchical relationships to mutual engagement, capturing the complexities of power dynamics and vulnerabilities. Other theoretical frameworks, such as patient-centered care, and concepts, such as patient empowerment, were also valuable for understanding these interactions in the context of digital health.

The shift from hierarchical to collaborative models in the physician-patient relationship not only underscores the empowering potential of digital tools but also presents new challenges and reinforces existing ones. Along with applications for various theoretical frameworks and concepts, this review highlights the unique comprehensiveness of a relational ethics perspective, which could provide a more nuanced understanding of trust, empathy, and power dynamics in the context of digital health. The adoption of relational ethics in empirical research may offer richer insights into the real-life complexities of the physician-patient relationship, as mediated by digital technologies.

KEYWORDS
digital health; mobile health; telemedicine; physician-patient relations; relational ethics; primary care; patient-provider; physician-patient; telehealth; relationship; eHealth; scoping review; review method; mobile phone

Introduction

Rationale
Digital health technologies are revolutionizing the practice of health care [1]. Driven by societal, political, and technological advancements, these technology-mediated changes are occurring at an unprecedented pace. Envisioning a future in which medicine can be personalized to individuals and diseases can be foreseen and prevented, some experts have high hopes for the potential of technology to improve human life [1,2]. However, skeptics argue that achieving personalization through technology is an elusive paradox, as it fails to account for the fundamental uncertainties and complexities of medicine and the value of human interactions [3].

In the realm of medical practice, a physician’s primary goal is to diagnose and treat patients’ diseases by drawing upon their biomedical expertise and years of clinical experience. Nevertheless, a physician’s practice must also be guided and adapted based on their relationship with the patient. A sustained physician-patient relationship facilitates the tailoring of therapeutic interventions to best suit individual patients’ needs [4]. Interpersonal skills honed over time through relational interactions with patients cannot be directly replaced by technology. Future technologies, including artificial intelligence, can only interpret what is explicitly documented as text or images in the electronic health record (EHR), which could lead to potentially missing valuable information generated within the physician-patient relationship, such as the patient’s values, preferences, trust, and rapport [5,6]. Moreover, a doctor’s personality and nonreflective actions, such as active listening or disregarding patient preferences, can have a significant impact not only on curing the patient’s illness but also on maintaining or exacerbating it [4]. Continuity of care with the same physician has been shown to reduce hospital admissions and mortality rates [7]. The rapid integration of technology introduces complex dimensions to an already intricate relationship, necessitating an exploration of technology’s impact on the physician-patient dynamic.

Digital technology offers potential for both improvements and risk, as it can efficiently reach a larger number of patients with less effort. At best, this technology-mediated efficiency may offer improved accessibility, real-time monitoring, and personalized treatments. It can also bridge geographical divides, connecting remote patients to specialists. At worst, it may be harmful on a scale far beyond what a single physician could achieve in a lifetime, potentially introducing systematic biases on a massive, automatic scale.

A potential problem with research on digital health technologies is the risk of bias, for example, when research funding comes from for-profit organizations and technology providers. As there may also be a dearth of independent research being published in this domain, we risk developing a skewed view of the evidence [8]. Additionally, while technological advancements surge ahead, ethical and regulatory frameworks have struggled to keep pace with these developments [9], making the need for new knowledge even more pressing.

Undeniably, digital technologies are here to stay and are already reshaping medical practice and the physician-patient relationship. Yet, more research is necessary to truly understand the nuanced implications of this shift. While technology promises efficiency and precision, the physician-patient relationship is rooted in intersubjective trust, empathy, and a deep understanding that transcends quantifiable data. As we further integrate technology into health care, it is therefore essential to explore its impact on this foundational relationship. Does it have the potential to augment the bond, creating new avenues for connection and understanding, or does it carry a more detrimental potential to create distance, becoming a screen that separates rather than connects human beings? By delving deeper into these questions, we can ensure that, as we advance technically, we do not lose sight of the human touch that remains at the heart of healing.

Objectives
This review explores the existing knowledge gap concerning the impact of digital health on the physician-patient dynamic. Additionally, we analyze various theoretical frameworks and concepts used in empirical studies concerning this relationship and contrast the findings of these studies with the results of research grounded in relational ethics frameworks [10]. This research focus has enabled us to identify the consequences of using digital health technologies (such as digital messaging, telemedicine, health-related websites, and EHRs) in physician-patient relationships.

Methods

Overview
Due to the rapid advancements in digital health care, there is a noticeable knowledge gap in the existing literature. A scoping review was deemed a suitable review approach, as it is well suited for providing an updated understanding of the current state of knowledge [11]. This scoping review was conducted according to the guidelines of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [12]. The study protocol was preregistered with the Center for Open Science [13].

Eligibility Criteria
The combined emergence of smartphones and 4G networks from around 2010 has played a crucial role in driving the development and use of digital health services [14]. This convergence has provided individuals with convenient access to health care services through mobile health (mHealth) apps, remote monitoring, and telemedicine, while also enabling health care providers to leverage real-time health data for personalized and data-driven care approaches. Only original peer-reviewed research published in English-language journals from January
1, 2010, to December 31, 2021, was included in this review. The decision to conclude the literature review on December 31, 2021, was driven by the aim to provide a comprehensive and current overview of the field up to our initial submission at the beginning of 2023. We carefully evaluated the scope and depth of available literature within this timeframe, determining that extending the search would likely not alter our core findings.

Research papers on patients and general practitioners or physicians involving health technology were included. Studies that did not specifically mention the physician-patient relationship were excluded.

Information Sources

Studies for inclusion were identified through searches in the following databases: MEDLINE/PubMed, Embase, IEEE Xplore, and Cochrane. Since our review specifically targeted empirical research on the physician-patient relationship, we primarily focused on searching medical databases. However, we also included IEEE Xplore to ensure that we did not miss any significant research in the fields of computer science and electrical engineering. To ensure trustworthy findings and direct insights, our search prioritized original papers and solely focused on high-quality research papers published in peer-reviewed medical journals. Reviews, opinion pieces, and other nonempirical texts were not included. Duplicates were removed.

Search

In collaboration with a librarian, we devised a comprehensive search string to explore the application of digital health technology within the physician-patient relationship. To ensure clarity in our scope, we used the mnemonic strategy PCC (population, concept, and context) to determine which papers should be included. Specifically, the population consisted of physicians, patients, or both; the concept involved digital health technologies used in medical care; and the context revolved around their implementation within the physician-patient relationship. The search string encompassed relevant terms, such as “digital health,” “mHealth,” “eHealth,” “telemedicine,” or “telehealth,” combined with the key phrases “physician-patient relations” or “relational ethics.” Detailed explanations of these terms are provided in Table 1. The search string used was (“digital health” OR “mHealth” OR “eHealth” OR “telemedicine” OR “telehealth”) AND (“physician-patient relations” OR “relational ethics”).

Our particular focus was on “relational ethics” due to its potential to inform a more ethical analysis of the physician-patient relationship, thus addressing the intricacies and challenges unique to this context.

We searched the 4 databases applying the following limitations: “scholarly (peer-reviewed) journals,” “date of publication from January 1, 2010 to December 31, 2021,” and “English language.” The search resulted in the identification of 429 journal papers. Of these papers, 427 were discovered in MEDLINE/PubMed, while 1 paper was identified as a duplicate appearing in both the Cochrane and MEDLINE/PubMed libraries. Additionally, 1 paper obtained from Embase was identified as a scoping review and was therefore excluded from our screening analysis. It is worth noting that no papers were found in IEEE Xplore during our search process. In short, all 427 papers selected for screening were found in the MEDLINE/PubMed database.

Table 1. Search terms.

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>digital health</td>
<td>Refers to technology related to health care services and solutions</td>
</tr>
<tr>
<td>mHealth</td>
<td>Short for “mobile health,” involves the use of mobile devices</td>
</tr>
<tr>
<td>eHealth</td>
<td>Refers to “electronic health,” including digital health services</td>
</tr>
<tr>
<td>telemedicine</td>
<td>Involves remote diagnosis and treatment using technology</td>
</tr>
<tr>
<td>telehealth</td>
<td>Broad term covering health care services through telecommunication</td>
</tr>
<tr>
<td>Physician-patient relations</td>
<td>Refers to the interactions and dynamics between doctors and patients</td>
</tr>
<tr>
<td>relational ethics</td>
<td>Refers to an ethical framework focused on understanding and navigating the intricacies of interpersonal dynamics in the doctor-patient relationship</td>
</tr>
</tbody>
</table>

*aHealth: mobile health.

Selection of Sources of Evidence

The screening process was performed by 2 researchers. It involved reviewing the abstracts initially found for inclusion, and in cases where abstracts lacked sufficient information, the entire research paper was examined. In instances where 2 researchers could not reach a consensus on whether to include a research paper, a third researcher was consulted as an arbiter to resolve any disagreements. This process led to the inclusion of 73 research papers. Subsequently, all 73 included papers underwent a detailed and thorough examination by 2 researchers. In the event of any disagreement during this phase, a third researcher was consulted to ensure accuracy. Studies that did not specifically mention the physician-patient relationship (n=63) were excluded, resulting in 10 studies to be reviewed.

Data Charting Process

A standardized data charting form was created for the data extraction process, and data charting was performed by 2 reviewers.

Data Items

Our objective was to collect a range of data items relevant to the research. These included the year of the study, the country...
where it was conducted, the study type (qualitative or quantitative), the theoretical or ethical framework and concepts used, the type of technology studied, the research objectives, and the participants involved (patients, physicians, and other health care personnel [HCP]; Table 2). Additionally, we extracted text samples that described the impact of digital health technologies on the physician-patient relationship. Our approach involved presenting the main findings in a textual manner.

**Table 2.** Overview of the data from the included studies.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Journal</th>
<th>Country</th>
<th>Study type</th>
<th>Type of technology</th>
<th>Aims</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balato et al [16]</td>
<td>2013</td>
<td>British Journal of Dermatology</td>
<td>Italy</td>
<td>Quantitative</td>
<td>Mobile phone messages</td>
<td>Evaluate the use of telemedicine in improving treatment adherence, patient outcomes, and the physician-patient relationship</td>
<td>Patients and physicians</td>
</tr>
<tr>
<td>Grünloh et al [17]</td>
<td>2018</td>
<td>Journal of Medical Internet Research</td>
<td>Sweden</td>
<td>Qualitative</td>
<td>Web-based patient portal</td>
<td>Investigate how physicians view the idea of patient participation</td>
<td>Physicians</td>
</tr>
<tr>
<td>Győrfy et al [18]</td>
<td>2020</td>
<td>PLOS ONE</td>
<td>Hungary</td>
<td>Qualitative</td>
<td>Social media</td>
<td>Explore physicians’ knowledge and attitudes toward digital health technologies and the transformation of the doctor-patient relationship</td>
<td>Physicians</td>
</tr>
<tr>
<td>Jiang [19]</td>
<td>2019</td>
<td>Health Communication</td>
<td>China</td>
<td>Quantitative</td>
<td>Digital messaging</td>
<td>Examine how the quality of face-to-face communication with providers is associated with their subsequent internet use for patient-provider communication</td>
<td>Patients</td>
</tr>
<tr>
<td>Kladacz-Alessandri et al [20]</td>
<td>2021</td>
<td>PLOS ONE</td>
<td>Poland</td>
<td>Qualitative</td>
<td>Teleconsultation (phone)</td>
<td>Study patients’ satisfaction with teleconsultation in primary care and the impact of teleconsultations on GP-patient communication</td>
<td>Patients and physicians</td>
</tr>
<tr>
<td>Macdonald et al [21]</td>
<td>2018</td>
<td>Journal of Medical Internet Research</td>
<td>Canada</td>
<td>Qualitative</td>
<td>Digital messaging and EHR</td>
<td>Examine HCP’s perspectives on how eHealth affects their relationships with patients, as well as its ethical ramifications</td>
<td>Patients, physicians, and other HCPs</td>
</tr>
<tr>
<td>Tasneem et al [22]</td>
<td>2019</td>
<td>American Journal of Hospice and Palliative Medicine</td>
<td>United States</td>
<td>Qualitative</td>
<td>Video consultation</td>
<td>Investigate the need for web-based videoconferences for oncology patients</td>
<td>Patients and physicians</td>
</tr>
<tr>
<td>Townsend et al [23]</td>
<td>2015</td>
<td>Journal of Medical Internet Research</td>
<td>Canada</td>
<td>Qualitative</td>
<td>Health-related websites</td>
<td>Focus on patients’ and HCP’s use of health-related internet information and how it influences the patient-HCP relationship</td>
<td>Patients, physicians, and other HCPs</td>
</tr>
</tbody>
</table>

\[a\]GP: general practitioner.
\[b\]EHR: electronic health record.
\[c\]HCP: health care personnel.

**Synthesis of Results**

To synthesize the results, we first created a summary based on the extracted data items to capture the key findings and statements from the included research papers concerning the physician-patient relationship. This summary was then compared with each research paper in its entirety to ensure accuracy. Next, we conducted a narrative synthesis of the findings from the included studies, incorporating both the extracted text samples and the summaries of the papers. This approach allowed us to provide a comprehensive overview of the results from both quantitative and qualitative studies, enabling a comprehensive summary of the physician-patient relationship in the context of digital health technologies (Table 3). In addition, we assessed whether the papers applied any ethical frameworks or simple concepts and analyzed how their use contributed to the research (Table 4).
The paper also mentions a pedagogical approach to interacting with the eHealth users among patients. One of the health care professionals stated that patients who are engaged through eHealth and informed about their health and illness are more useful clinically. This shift toward more collaborative interactions with patients is, in part, facilitated by eHealth technologies. The authors mention the importance of patient participation for patient safety. The role of the doctor is in transition, with doctors expected to perform more complex tasks including health information technology and aiding in the digital orientation of patients. They see they transforming into mediators based on efficient communication with their patients. Digitally engaged physicians consider themselves guides, undertaking a guardian and information managing function in the description, collection, and sharing of credible content in the online space. For a successful leap from hierarchical patterns to the 21st-century doctor-patient relationship, the future generation of physicians should be trained differently and prepared for all the above-described changes. Medical school curricula should emphasize health and prevention rather than only diseases and pathology via the newest digital technological solutions. Medical students need to prepare for predictive and proactive working environments, including their new role as a guide or mediator for digitally empowered patients, in contrast with the paternalistic physicians of previous generations.

Summary of evidence:

<table>
<thead>
<tr>
<th>Authors</th>
<th>Summary: What does the paper state about the physician-patient relationship?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audrain-Pontevia and Menielle</td>
<td>• The paper discusses the impact of online health communities on the patient-physician relationship. The authors explore how online health communities, which provide users with computer-mediated social support and empowerment, impact this relationship.</td>
</tr>
<tr>
<td></td>
<td>• The authors acknowledge that, traditionally, doctors were the main source of medical information and therefore benefited from authority and power over their patients. However, with the advent of online health communities, patients now have access to social support, resources, and aid, which can make them feel more empowered and influence their relationships with their physicians.</td>
</tr>
<tr>
<td></td>
<td>• The authors propose that online health communities offer patients the opportunity to gain the power to handle their illnesses and their health, presumably leading to increased participation during the consultation and improving their commitment to their relationship with their physician.</td>
</tr>
<tr>
<td>Balato et al [16]</td>
<td>• The study found that patient-physician communication improved in a group receiving SMS text message interventions, whereas it remained unchanged in the control group. This suggests that the use of digital interventions, such as SMS text messages, could potentially enhance the patient-physician relationship.</td>
</tr>
<tr>
<td>Grünloh et al [17]</td>
<td>• The models of the doctor-patient relationship presented in the paper describe patients as being static and unchanged, but the authors note that patients with chronic conditions often encounter new situations and need to engage in a sensemaking and learning process.</td>
</tr>
<tr>
<td></td>
<td>• The paper suggests that the use of patient-accessible EHRs can contribute to the development of the doctor-patient relationship by allowing patients to play an active role. This increased patient participation makes it more difficult for physicians to maintain a strategy that potentially excludes patients.</td>
</tr>
<tr>
<td></td>
<td>• The authors state that eHealth does not have to be a “power struggle” in the doctor-patient relationship but can potentially help both partners improve their relationship collectively and grow individually.</td>
</tr>
<tr>
<td></td>
<td>• The authors mention the importance of patient participation for patient safety.</td>
</tr>
<tr>
<td>Győrffy et al [18]</td>
<td>• The role of the doctor is in transition, with doctors expected to perform more complex tasks including health information technology and aiding in the digital orientation of patients. They see themselves transforming into mediators based on efficient communication with their patients.</td>
</tr>
<tr>
<td></td>
<td>• Digitally engaged physicians consider themselves guides, undertaking a guardian and information managing function in the description, collection, and sharing of credible content in the online space.</td>
</tr>
<tr>
<td></td>
<td>• For a successful leap from hierarchical patterns to the 21st-century doctor-patient relationship, the future generation of physicians should be trained differently and prepared for all the above-described changes.</td>
</tr>
<tr>
<td></td>
<td>• Medical school curricula should emphasize health and prevention rather than only diseases and pathology via the newest digital technological solutions.</td>
</tr>
<tr>
<td></td>
<td>• Medical students need to prepare for predictive and proactive working environments, including their new role as a guide or mediator for digitally empowered patients, in contrast with the paternalistic physicians of previous generations.</td>
</tr>
<tr>
<td>Jiang [19]</td>
<td>• The results of the study emphasize the important roles of patient-centered communication and the physician-patient relationship in the eHealth and mHealth movement, particularly in the Chinese health care system.</td>
</tr>
<tr>
<td></td>
<td>• The interplay of physician-patient communication in face-to-face environments and relationship factors (eg, patient trust and patient satisfaction) could exert significant effects in promoting eHealth adoption.</td>
</tr>
<tr>
<td></td>
<td>• To encourage patients to adopt eHealth technologies, health care providers should first build a patient-centered environment (eg, responding to patients’ informational and emotional needs and engaging patients in medical decision-making).</td>
</tr>
<tr>
<td>Kladacz-Alessandri et al [20]</td>
<td>• The study concerns patients’ satisfaction with teleconsultation in primary care and the impact of teleconsultations on GP-patient communication during the COVID-19 pandemic in Poland. The paper suggests that the quality of GP-patient communication is an essential factor that can improve the results of treatment and patient satisfaction. Only 55% (n=99) of the patients found teleconsultation to be as good as in-person visits with their physician.</td>
</tr>
<tr>
<td>Macdonald et al [21]</td>
<td>• The study discusses the concept of a “two-way conversation” that is evolving in the health care provider-patient dynamic. This shift toward more collaborative interactions with patients is, in part, facilitated by eHealth technologies. The authors examine the impact of eHealth on the current state of collaborative consultation, highlighting how it aids in having, using, and supporting conversations with patients.</td>
</tr>
<tr>
<td></td>
<td>• Some health care professionals in the study embraced the idea of patients as “partners,” as they see a partner as someone who helps in improving an outcome by educating themselves and conscientiously monitoring their condition and behavior.</td>
</tr>
<tr>
<td></td>
<td>• One of the health care professionals stated that patients who are engaged through eHealth and informed about their condition are more useful clinically.</td>
</tr>
</tbody>
</table>
|                                  | • The paper also mentions a pedagogical approach to interacting with the eHealth users among patients.
The study assessed the needs of patients receiving palliative care and their perception of how telemedicine video visits might influence their care. Despite concerns about truncated physical examinations and prescription limits, the majority of patients favored having the opportunity for telemedicine video visits. They felt that the physician-patient relationship would not diminish and had few cost concerns. They believed that a video alternative to an in-person visit might increase access, save time, and increase comfort and safety by avoiding a trip to the hospital.

In this study, patients reported using personal websites, blogs, chat rooms, and online links to medical test results as part of their eHealth resources. This suggests that patients are actively engaging with digital resources in managing their health, which can have implications for the physician-patient relationship. The paper discusses how the rapid explosion in online digital health resources is seen as transformational, accelerating the shift from traditionally passive patients to patients as partners. This is altering the current patient-health care professional relationship.

The proliferation of eHealth strategies is accelerating a shift in health care from a traditional and paternalistic delivery model to a more mutual patient-health care professional relationship in which informed patients are actively involved in their care and treatment decisions. The authors mention that eHealth resources provide patients with extensive and up-to-date information, access to medical research, connections to people with similar conditions, immediacy, and convenience in patient-health care professional communications. These factors can significantly impact the physician-patient relationship.

The paper investigates the underlying reasons for poor doctor-patient relationships in mobile consultation from the perspective of computer-mediated communication. This suggests that the physician-patient relationship may be influenced by the mode of communication, particularly in a digital context. The paper emphasizes the emerging use of mobile medical consultation in China, which has propelled the establishment of doctor-patient relationships in the mobile context. This again underscores the impact of digital technologies on the physician-patient relationship.

The authors also mention different models or concepts used to assess the doctor-patient relationship, suggesting that the nature of this relationship can be complex and multifaceted.

---

**Table 4. Theoretical frameworks and concepts used in the studies.**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Framework or concepts</th>
<th>Relevance of using specific theoretical frameworks and concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balato et al [16]</td>
<td>None mentioned</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Grünloh et al [17]</td>
<td>Shared decision-making, patient-centered care, and paternalism</td>
<td>Assesses the impact of a web-based patient portal on patient participation.</td>
</tr>
<tr>
<td>Györffy et al [18]</td>
<td>Patient empowerment and shared decision-making</td>
<td>Studies the attitudes of digitally engaged physicians toward transforming the physician-patient relationship.</td>
</tr>
<tr>
<td>Jiang [19]</td>
<td>Patient-centered care, mutual trust, and patient satisfaction</td>
<td>Examines the relationship between face-to-face and online patient-provider communications.</td>
</tr>
<tr>
<td>Macdonald et al [21]</td>
<td>Relational ethics</td>
<td>Relational ethics addresses the ethical content and decisions implicit in everyday relationships and conversations.</td>
</tr>
<tr>
<td>Tasneem et al [22]</td>
<td>None mentioned</td>
<td>N/A</td>
</tr>
<tr>
<td>Townsend et al [23]</td>
<td>Relational ethics</td>
<td>Core elements of relational ethics are applicable to everyday experiences, practice, and interactions. Applying relational ethics helps with focusing on what is valued in interactions and relationships and what is at risk rather than specific aspects of eHealth such as the nature of self-monitoring devices.</td>
</tr>
<tr>
<td>Yan et al [24]</td>
<td>None mentioned</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup>EHR: electronic health record.<br>
<sup>b</sup>mHealth: mobile health.<br>
<sup>c</sup>GP: general practitioner.
Results

Selection of Sources of Evidence

Out of the 427 studies screened, 73 studies were sought out for retrieval, and after excluding 63 studies that did not focus on the physician-patient relationship, 10 studies were included in this review. Figure 1 illustrates the search process.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) diagram: papers included in the review.

Characteristics of Sources of Evidence

The research papers included in this study [15-24] were conducted in various countries, with 3 in Canada, 2 in China, 1 in Italy, 1 in Sweden, 1 in Hungary, 1 in Poland, and 1 in the United States. These papers were published between 2013 and 2021, with half of them published in 2019 or more recently. The majority of the papers (n=7) were qualitative studies, with the remaining (n=3) papers being quantitative studies. Only 2 research papers, authored by Macdonald et al [21] and Townsend et al [23], referred to a specific ethical framework. Both used relational ethics. A total of 6 other papers adopted different theoretical frameworks and concepts, while the remaining 2 did not use any theoretical frameworks at all.

Regarding the technologies investigated, 3 papers focused on online communities, 3 papers focused on messaging, 2 papers focused on telephone or video consultation, and 2 papers focused on patient portals or EHR. The majority of the papers (n=6) included a mix of patients and physicians or other health care professionals, while 2 papers focused solely on patients and 2 others exclusively on physicians.

For further details about the characteristics of the included research papers, please refer to Table 2.

The Impact of Digital Health Technologies on the Physician-Patient Relationship

Kludacz-Alessandri et al [20] noted that, while teleconsultations offer convenience and safety, especially in situations like the COVID-19 pandemic, they may not always provide the same level of patient satisfaction as in-person visits. According to the authors, this could be due to several reasons such as the inability to read nonverbal cues, technical difficulties, or the impersonal nature of digital communication. The authors suggested that, while digital technologies offer new avenues for communication, they may not fully mirror the richness and depth of face-to-face interactions.

Yan et al [24] discussed the use of mobile medical consultation in China and noted the potential challenges in establishing effective physician-patient relationships in a digital context. According to the authors, 1 key challenge could be the varying degrees of technology literacy among patients. Not all patients may be comfortable using digital platforms for health care, and some may lack access to the necessary technology. This study indicates that this could lead to inequities in access to health care services and affect the quality of the physician-patient relationship.

When patients become more empowered through access to digital health resources and online communities, changes could occur in the power dynamics of the physician-patient relationship.
relationship. Some physicians may struggle to adapt to these changes, particularly if they are trained in a more traditional, paternalistic model of health care. This was suggested in the study by Audrain-Pontevia and Menvielle [15].

The advent of online health communities and access to a wealth of health-related information on the internet has empowered patients to take an active role in managing their health. As discussed by Audrain-Pontevia and Menvielle [15], this empowerment can lead to increased participation during consultations and improve commitment to the relationship with the physician.

Digital technologies, such as SMS text messaging, teleconsultations, and mHealth apps, can enhance communication between patients and physicians, making it more frequent, timely, and convenient. As pointed out by Balato et al [16], such digital interventions can potentially enhance the patient-physician relationship.

The use of eHealth and mHealth technologies facilitates a more patient-centered approach to care. As Jiang [19] noted, health care providers who build a patient-centered environment—responding to patients’ informational and emotional needs and engaging them in medical decision-making—can promote eHealth adoption and improve health care outcomes.

Digital health technologies can improve the efficiency of health care delivery and make health services more accessible. For example, Tasneem et al [22] found that patients receiving palliative care favored the opportunity for video consultations, as they could save time while increasing access, comfort, and safety by avoiding a trip to the hospital.

eHealth resources provide patients with extensive and up-to-date information, access to medical research, and connections to people with similar conditions. Townsend et al [23] mentioned that these factors can significantly impact the physician-patient relationship, transforming patients from passive recipients of care into active partners in their health care journey.

The availability of health information online and through digital technologies enables patients to be more informed about their health conditions and treatment options. This can lead to better-shared decision-making between patients and physicians [17,18,21].

The Application of Theoretical Frameworks and Concepts

In the study by Grünloeh et al [17], the use of 2 theoretical frameworks (shared decision-making and patient-centered care) and 1 concept of the physician-patient relationship (paternalism) helped highlight the roles of the medical professional and the patient, as well as the ways in which patients can contribute to the relationship.

The concept of patient empowerment (used in the study by Audrain-Pontevia and Menvielle [15]) helped with understanding how online health communities influence the patient-physician relationship, but it did not capture the complexities of the power dynamics and relational aspects inherent in these interactions.

Györffy et al [18] focused on the attitudes of digitally engaged physicians toward transforming the physician-patient relationship. The researchers used the theoretical frameworks of patient empowerment and physician-patient collaboration to assess current and ideal physician-patient relationships. Through this, they recognized that the digital age requires physicians to transition from a role of authority to one of guidance.

Jiang [19] examined the relationship between face-to-face and online patient-provider communication through the concepts of patient-centered care, trust, and satisfaction.

Kludacz-Alessandri et al [20] investigated patients’ satisfaction with teleconsultations by considering the concepts of respect and dialogue. These concepts are indeed valuable in understanding the impacts of teleconsultations on the physician-patient relationship.

The 3 studies that did not apply any theoretical frameworks had either superficial findings regarding the physician-patient relationship or it was not clear how they reached their conclusions. For example, Balato et al [16] simply quantified the physician-patient relationship through a 10-point scale questionnaire. Tasneem et al [22] asked patients only about how the technology would affect their relationship with their physician in 2 out of 15 questions. Finally, Yan et al [24] seemed to apply the concepts of mutual respect and patient satisfaction in a deep and meaningful way, but we have no way of knowing since they do not specifically state what concepts they used.

The Application of Relational Ethics

In contrast with the above-applied frameworks, the use of a relational ethics framework can help reveal the subtle and complex impacts of digital health technologies on the physician-patient relationship. For example, in the study by Macdonald et al [21], a relational ethics lens was used to examine how eHealth technology contributes to changes in relations between HCPs and patients, evolving toward more collaborative care. By focusing on every day relationships and conversations, the study was able to understand how these technologies incorporated the relational ethics of patient-centered care into practice.

Relational ethics can help address the power dynamics and vulnerabilities that come into play in the physician-patient relationship, especially with the use of digital technologies. In the study by Townsend et al [23], relational ethics was used to understand how technology impacts relational shifts in ethical patient-HCP relationships. They found that technology use could lead to a transition from a traditional hierarchical relationship to a more reciprocal relationship, which could reveal mutual vulnerabilities.

Discussion

Principal Results

The studies included in this review were diverse, with research conducted across various countries and contexts and exploring a range of technologies, such as online communities, messaging, teleconsultations, and patient portals or EHRs. In terms of user groups, they examined the experiences of patients, physicians,
and other HCPs, offering a comprehensive view of the phenomenon from different user perspectives.

Through the review, we identified several consequences of using digital health technologies in health care. The convenience and accessibility offered by these technologies have the potential to transform the health care landscape by enhancing communication, improving efficiency, and empowering patients. However, they also provide challenges and complexities such as increasing inequities in access to health care services due to variations in technology literacy among patients and decreases in patient satisfaction due to limitations in nonverbal communication and the impersonal nature of digital interactions.

The review further highlighted the shift in power dynamics in the physician-patient relationship from a traditional hierarchical model toward a more reciprocal and collaborative model. This shift is facilitated by the empowering potential of digital health technologies and online health communities but may also present challenges for physicians trained in a more paternalistic model of approach to medicine.

In addition to mapping key findings, we analyzed the theoretical frameworks used in the studies to contrast the use of relational ethics with the application of other theoretical frameworks. Of the 10 studies in this review, 2 did not apply any theoretical frameworks to their research. Others used frameworks or concepts such as patient-centered care, patient empowerment, or shared decision-making. Only 2 studies used relational ethics as their framework.

Relational ethics, by its very nature, emphasizes the value of intersubjective qualities such as empathy, trust, respect, and mutual responsibility. These qualities are critical to the practice of effective and compassionate health care, and their importance is highlighted in the context of digital health technologies interjected between the physician and the patient. While theoretical frameworks and concepts, such as patient-centered care and patient empowerment, provided valuable insights into the impacts of digital health technologies, a relational ethics perspective could provide a more comprehensive understanding. For instance, it could explore how trust is built and maintained in online versus face-to-face interactions, how the power dynamics between physicians and patients might shift in an online context, and how empathy and understanding are conveyed through digital mediums. A relational ethics framework could further enrich this understanding by examining how respect and dialogue contribute to a sense of mutual trust and understanding, how they shape the power dynamics in a teleconsultation, and how they can foster a sense of connection and empathy in a digital environment.

Comparison to Prior Work

This review underscores the potential value of using a relational ethics framework in research on the physician-patient relationship in the complex real-life context of digital health technologies. While previous works have acknowledged the importance of relational aspects in health care delivery [3,5], few have explicitly used relational ethics as a framework to examine the nuanced ethical implications of digital health technologies on the physician-patient relationship.

Limitations and Strengths of This Study

Because this review amassed studies from a variety of countries, we acknowledge that the specific cultural, societal, and health care contexts of these regions can have a significant effect on the physician-patient relationship and the use and adoption of digital health technologies. As such, our findings might not be applicable everywhere.

The review only included studies published in English, which could introduce language bias. If there were relevant studies published in other languages, they would have been left out, potentially narrowing the range of perspectives and findings we were able to consider.

While our review touched on a selection of digital health technologies—including online communities, messaging services, teleconsultations, and patient portals or EHRs—many more kinds of technology are being used in health care. As such, our findings are limited to the specific digital health technologies covered in this review.

Scoping reviews, like this one, are primarily intended to provide a broad overview of the existing literature rather than evaluate the strength of evidence or perform a meta-analysis. This means that it is difficult to draw definitive conclusions or provide firm recommendations based on our findings.

While our review encapsulates literature up to the end of 2021, reflecting the state of research at the time of our initial submission, it may not include the latest developments or studies published post-2021.

The quality of the studies included in a review can have a large effect on its findings. Since we did not conduct a critical appraisal of the quality of the studies we included, this could be seen as a limitation.

Conclusions

Overall, this review offers a comprehensive overview of the current state of evidence regarding the impacts of digital health technologies on the physician-patient relationship. It underscores the potential of these technologies to transform health care delivery, while also highlighting the challenges and complexities they introduce. The review emphasizes the need for further research using a relevant ethics framework to provide a deeper understanding of the impact of digital health technologies on the physician-patient relationship. This will be particularly crucial as digital health technologies continue to expand, evolve, and become more integrated into health care delivery.
Acknowledgments

The authors would like to thank librarian Grete Mortensen at the University of Stavanger for her support in developing the search strategy. The study was conducted as part of the research project “Caring Futures: Developing Care Ethics for Technology-Mediated Care Practices” (QUALITECH). QUALITECH was funded by the Research Council of Norway (grant NFR 301827).

Authors’ Contributions

DN was the principal author and was involved in all aspects of the scoping review. BHG was a coauthor and screened the abstracts along with DN. ER also coauthored this paper, as well as assessed the research papers for eligibility and reviewed the included research papers along with DN.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. [PDF File (Adobe PDF File), 547 KB - jmr_v13i1e47280_app1.pdf]

References


Abbreviations

- **EHR:** electronic health record
- **HCP:** health care personnel
- **mHealth:** mobile health
- **PCC:** population, concept, and context
- **PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Review

©Damoun Nassehi, Birgitta Haga Gripsrud, Ellen Ramvi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 15.05.2024. 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.

https://www.i-jmr.org/2024/1/e47280
Evaluation of an mHealth App on Self-Management of Osteoporosis: Prospective Survey Study

Magnus Grønlund Bendtsen1, MSc; Bodil Marie Thuesen Schönwandt1, MSc; Mette Rubæk1,2, MSc; Mette Friberg Hitz1,3, PhD

1Research Unit, Medical Department, Zealand University Hospital, Koege, Denmark
2University of Southern Denmark, Odense, Denmark
3Institute of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Magnus Grønlund Bendtsen, MSc
Research Unit, Medical Department, Zealand University Hospital
Lykkebækvej 1
Koege, 4600
Denmark
Phone: 45 22893352
Email: mag.bendtsen@gmail.com

Abstract

Background: Mobile health (mHealth) technologies can be used for disease-specific self-management, and these technologies are experiencing rapid growth in the health care industry. They use mobile devices, specifically smartphone apps, to enhance and support medical and public health practices. In chronic disease management, the use of apps in the realm of mHealth holds the potential to improve health outcomes. This is also true for mHealth apps on osteoporosis, but the usage and patients’ experiences with these apps are underexplored.

Objective: This prospective survey study aimed to investigate the eHealth literacy of Danish patients with osteoporosis, as well as the usability and acceptability of the app “My Bones.”

Methods: Data on patient characteristics, disease knowledge, eHealth literacy, usability, and acceptability were collected using self-administered questionnaires at baseline, 2 months, and 6 months. The following validated questionnaires were used: eHealth Literacy Questionnaire, System Usability Scale, and Service User Technology Acceptability Questionnaire.

Results: Mean scores for eHealth literacy ranged from 2.6 to 3.1, with SD ranging from 0.5 to 0.6 across the 7 domains. The mean (SD) System Usability Scale score was 74.7 (14.4), and the mean (SD) scores for domains 1, 2, and 6 of the Service User Technology Acceptability Questionnaire were 3.4 (1.2), 3.5 (1.1), 4.1 (1.2), respectively.

Conclusions: Danish patients with osteoporosis are both motivated and capable of using digital health services. The app’s usability was acceptable, and it has the potential to reduce visits to general practitioner clinics, enhance health outcomes, and serve as a valuable addition to regular health or social care services.

doi:10.2196/53995

KEYWORDS
eHealth literacy; health literacy; mHealth; mobile health; eHealth; mobile health apps; self-management; osteoporosis; usability; acceptability

Introduction

Background

Osteoporosis ranks as the fourth most prevalent chronic disease globally, carrying substantial negative personal and economic consequences [1]. Osteoporosis affects approximately 40% of women and 17% of men aged 50 years or older. Despite its high prevalence, osteoporosis is significantly underdiagnosed [2]. Although osteoporosis may remain asymptomatic for many individuals, it presents significant risks such as fractures, chronic pain, reduced daily activity, compromised quality of life, and increased mortality [1]. Lack of disease-specific knowledge among patients with osteoporosis is a global issue [3]. It has been shown that improved support for patients to understand
osteoporosis upon diagnosis is required, along with support in self-management of the disease [4,5].

Mobile health (mHealth) technologies can be used for disease-specific self-management, and the technologies are experiencing rapid growth in the health care industry. These technologies use mobile devices, specifically smartphone apps, to enhance and support medical and public health practices [6]. Studies have demonstrated that, particularly in chronic disease management, the use of apps in the realm of mHealth holds the potential to enhance health outcomes [7]. Despite Slomian et al [8] addressing the potential of using mHealth apps to aid patients in self-managing osteoporosis in 2014, the implementation of such apps in the field of osteoporosis remains inadequate. A recently published systematic review and meta-analysis of digital health technologies for long-term self-management of osteoporosis identified 23 relevant apps for osteoporosis self-management and concluded that osteoporosis apps have the potential to support and improve the management of the disease. Furthermore, mHealth osteoporosis apps also appear to be valuable tools for patients and health care professionals. However, most of the identified apps that are currently available in the field of osteoporosis lack clinically validated evidence of their efficacy [9].

From 2015 to 2018, a team of researchers and health care professionals from Odense University Hospital, Denmark, developed and tested an mHealth app for women recently diagnosed with osteoporosis [10]. The development of the app was based on identified needs among patients and health care professionals [11,12]. The app evaluation revealed that patients perceived the app as providing confidence and reassurance, fostering equitable dialog during consultations, and offering readily accessible assistance for self-managing osteoporosis [13]. After the test period, the app was implemented at Odense University Hospital, Denmark. In 2019, the Danish Health Authority decided to support a nationwide rollout of the app, named “My Bones.” A group of health care professionals with expertise in osteoporosis were engaged in the app’s further development and testing. In September 2021, the app “My Bones” was launched as a freely available app on both the App Store and Google Play platforms. Visuals of the app are available in Multimedia Appendix 1.

After the app’s launch, the Research Unit of the Medical Department at Zealnd University Hospital, Denmark, initiated an evaluation of the app focusing on usability, acceptability, eHealth literacy, and self-perceived knowledge of osteoporosis. The Consumer Health Information System Adoption Model was used as a theoretical framework, as a basis for the choice of questionnaires used in this study. In this model, eHealth literacy of the user, usefulness, and usability are all determining factors for the adoption of consumer health information systems like the app “My Bones.”

Objective
The aim of this study was to investigate the eHealth literacy of Danish patients with osteoporosis and to assess the usability and acceptability of the app “My Bones.” Additionally, the aim was to assess the current level of disease knowledge within the patient group.

Methods
Study Design
This prospective survey study was conducted at Zealand University Hospital, Denmark, from January 2020 to February 2023. A total of 100 patients were recruited to test the app “My Bones.” At the time of the study, the app was owned and operated by OSAIA Health. The app provided patients with comprehensive information about osteoporosis, its risk factors, available treatment options, the diagnostic procedure for osteoporosis, and the process of a dual-energy x-ray absorptiometry (DXA) scan. Alongside this general information, the app provided guidance on maintaining a bone-healthy lifestyle, dietary requirements including calcium and vitamin D supplements, and recommendations for physical activity. App users also gained access to basic and safe training exercises tailored to their individual fracture risk and functional level, enabling them to engage in a foundational level of physical activity. The app includes 2 interactive modules. One of these was a calcium calculator that assists patients in measuring their daily calcium intake and determining whether they require calcium supplements. The second module was an interactive DXA graph that gives patients the option to plot their T score values for the spine and hip from each DXA scan.

Three self-administered questionnaires were developed and internally pretested before distribution at baseline (Q0), at 2 months (Q2), and at 6 months (Q6). The questionnaires are described under the Measures section. The results from the baseline (Q0) and 2-month (Q2) questionnaires are presented here. However, the results from the 6-month (Q6) questionnaire are excluded because of technical issues preventing data analysis.

Sampling and Recruitment
Patients were recruited from the Outpatient Clinic of Endocrinology at Zealand University Hospital and through advertisements on the Danish Osteoporosis Association’s homepage and a Facebook page.

The inclusion criteria include patients diagnosed with osteoporosis based on either the T score criterion (T ≤ –2.5 at the lumbar spine, total hip, or hip neck) or a diagnostic osteoporotic fracture (fragility fracture of a vertebra with >20% compression or a hip fracture). Postmenopausal women or men over 45 years of age who can read and understand Danish and have access to, as well as the ability to use, a smartphone, tablet, or computer.

The exclusion criteria include patients with mental and cognitive conditions impairing their ability to use an app and read and understand questionnaire questions.

Procedures
Patients with osteoporosis interested in participating were provided with both written patient-oriented materials and oral information about the study. Informed consent was obtained from the patients by the data controller. The signed consent forms were stored in a secure location, only accessible by the data controller. After signing the informed consent, participants were asked to indicate their preferred method of receiving the
study questionnaires—either by email, through e-Boks (a digital postbox for communication between companies, public authorities, and private citizens), or in a printed copy. Detailed instructions on how to download the app from the App Store or Google Play were provided. If patients required assistance, the study staff aided in the process.

SurveyXact, a tool for creating electronic questionnaire-based surveys, was used to distribute the questionnaires to the patients and to establish a database. The data analyst gained access to the database only after all data had been collected and anonymized by way of a respondent key in the survey tool and transferred to the database. The survey tool is compliant with the European Union’s General Data Protection Regulation.

In cases of nonresponse, electronic reminders were dispatched, followed by telephone follow-ups. If participants still did not complete the questionnaire despite these attempts, they were classified as dropouts.

Measures

**Baseline Questionnaire (Q0)**

Sociodemographic data include age, sex, education, occupation, time since diagnosis, and knowledge of the disease (self-reported T score of the spine and hip, medical treatment, and fracture history).

Data regarding education were initially reported on 6 levels and were subsequently consolidated into two levels for statistical analysis: (1) shorter education and (2) longer education. Longer education was defined as a bachelor’s degree or higher. Data concerning occupation were reported across 8 levels and were later condensed into two levels for statistical analysis: (1) currently working and (2) not currently working. The various education and occupation levels can be found in Table 1.

The eHealth Literacy Questionnaire (eHLQ) consists of 35 items representing 7 scales covering the eHealth Literacy Framework dimensions. Each scale has 4 to 6 items with a 4-point response option. Mean scores are calculated from each scale with equal weighting [14].

The International Physical Activity Questionnaire measures the amount of time an individual spends physically active during a normal 7-day week [15].

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>66.3 (8.4)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5 (5)</td>
</tr>
<tr>
<td>Female</td>
<td>90 (95)</td>
</tr>
<tr>
<td><strong>Shorter education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>42 (44)</td>
</tr>
<tr>
<td>Vocational training, high school</td>
<td>21 (22)</td>
</tr>
<tr>
<td>Higher education—short</td>
<td>12 (14)</td>
</tr>
<tr>
<td><strong>Longer education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Higher education—intermediate</td>
<td>53 (56)</td>
</tr>
<tr>
<td>Higher education—long</td>
<td>41 (43)</td>
</tr>
<tr>
<td><strong>Currently working, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Self-employed (professional)</td>
<td>3 (3)</td>
</tr>
<tr>
<td>Civil servant</td>
<td>23 (24)</td>
</tr>
<tr>
<td>Vocational</td>
<td>3 (3)</td>
</tr>
<tr>
<td>Unskilled or semiskilled worker</td>
<td>4 (4)</td>
</tr>
<tr>
<td><strong>Not currently working, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>62 (65)</td>
</tr>
<tr>
<td>Early retirement</td>
<td>17 (18)</td>
</tr>
<tr>
<td>Retirement pension</td>
<td>41 (43)</td>
</tr>
<tr>
<td>Leave of absence</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>
Two-Month Questionnaire (Q2)

The System Usability Scale (SUS) is a 10-item Likert scale questionnaire that offers a global perspective on subjective assessments of usability. The SUS generates a single number that represents a composite measure of the overall usability of the system under study, in this case, an app [16-18]. The questionnaire is validated in the Danish population.

The Service User Technology Acceptability Questionnaire (SUTAQ) is a questionnaire that consists of 22 items divided into 6 different subscales. For the purpose of this study, 3 of the 6 domains were included in the Q2 questionnaire. The included domains are (1) “Enhanced Care,” (2) “Increased Accessibility,” and (3) “Satisfaction.” The sum of each subscale indicates the degree of average internal agreement with it [19]. The domains “Privacy and Discomfort,” “Care Personnel Concerns,” and “Kit as a Substitution” were excluded because of their irrelevance resulting from the lack of health monitoring and interaction between patients and health care professionals via the app. The questionnaire is validated in the Danish population.

Statistical Analysis

Statistical analysis was conducted using the SPSS statistical software (version 21; IBM Corp). Descriptive statistics were generated for participants’ characteristics and other dependent variables, with calculation of mean values and SD for normally distributed data and median with range for data not normally distributed. Frequency data were calculated for categorical data.

The relation between each of the 7 eHLQ domains and the covariates age, education, and occupation was investigated using a backward stepwise linear regression analysis.

The correlation between the SUS questionnaire score and age was investigated using the Pearson 2-tailed correlation analysis.

All statistical outcomes were examined against a P value of .05 to determine statistical significance. The study was not an intervention study, and no power calculation was performed.

Ethical Considerations

The study was conducted in accordance with the principles of the Declaration of Helsinki. It received approval from the local Data Protection Authority (Reg-152-2020). Although the study did not meet the criteria that necessitate approval from the ethics committee, guidelines for obtaining informed consent were adhered to.

Results

Participant Characteristics

A total of 100 patients signed the informed consent. Of these, 95 responded to the entire or part of the baseline questionnaire (Q0), as shown in the flow diagram in Figure 1.

Of the 95 respondents, 90 were women and 5 were men. The mean (SD) age of the study population was 66.3 (8.4) years, and education levels were reported as shorter by 44% (42/95) of the participants and as longer by 56% (53/95) of the respondents. Regarding occupational status, 35% (33/95) stated a current connection to the labor market, while 65% (62/95) stated that they were either retired, unemployed, or on leave of absence (Table 1).

The mean time since the diagnosis of osteoporosis was 6 years (0-34 years). Among the 95 respondents, 52% (49/95) did not know the T score of the lumbar spine, and 68% (64/95) did not know the T score of the hip. Regarding self-reported osteoporosis status, 22% (21/95) and 4% (4/95) reported severe osteoporosis (T score<–3.0) in the spine region and the hip region, respectively (Table 2). Previous major osteoporotic fractures were reported by 49% (46/95) of the respondents (Table 2), while 67% (64/95) stated that they were currently undergoing medical treatment, the majority in treatment with bisphosphonate (Table 3).

Too much missing data hindered a meaningful analysis of the data from the International Physical Activity Questionnaire, and as a consequence, the results cannot be presented.
**Figure 1.** Participant flow diagram.

100 participants signed the informed consent form

1 dropped out after informed consent

99 participants initiated Q0

99 participants completed all of Q0

5 participants completed part of Q0

73 participants completed all of Q2

7 participants completed part of Q2

Table 2. *T* scores and fractures (N=95).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time since diagnosis(years), mean (range)</td>
<td>6 (0-34)</td>
</tr>
<tr>
<td><em>T</em> score lumbar spine, n (%)</td>
<td></td>
</tr>
<tr>
<td>-2.5 to -3.0</td>
<td>24 (26)</td>
</tr>
<tr>
<td>-3.0 to -4.0</td>
<td>14 (15)</td>
</tr>
<tr>
<td>≥-4.0</td>
<td>7 (7)</td>
</tr>
<tr>
<td>Do not know my <em>T</em> score</td>
<td>49 (52)</td>
</tr>
<tr>
<td><em>T</em> score hip, n (%)</td>
<td></td>
</tr>
<tr>
<td>-2.5 to -3.0</td>
<td>26 (28)</td>
</tr>
<tr>
<td>-3.0 to -4.0</td>
<td>3 (3)</td>
</tr>
<tr>
<td>≥-4.0</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Do not know my <em>T</em> score</td>
<td>64 (68)</td>
</tr>
<tr>
<td>Previous fractures, n (%)</td>
<td>46 (49)</td>
</tr>
<tr>
<td>Wrist</td>
<td>21 (22)</td>
</tr>
<tr>
<td>Upper arm</td>
<td>3 (3)</td>
</tr>
<tr>
<td>Vertebrae</td>
<td>15 (16)</td>
</tr>
<tr>
<td>Hip</td>
<td>6 (6)</td>
</tr>
<tr>
<td>Othera</td>
<td>19 (20)</td>
</tr>
</tbody>
</table>

*a* Other fractures reported: fracture of foot, ankle, heel, toe, finger, elbow, and tibia.
Table 3. Medication (N=93).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current treatment plan, n (%)</td>
<td></td>
</tr>
<tr>
<td>Oral BP&lt;sup&gt;a&lt;/sup&gt;</td>
<td>16 (25)</td>
</tr>
<tr>
<td>IV&lt;sup&gt;b&lt;/sup&gt; BP</td>
<td>26 (41)</td>
</tr>
<tr>
<td>Denosumab</td>
<td>11 (17)</td>
</tr>
<tr>
<td>PTH&lt;sup&gt;c&lt;/sup&gt; analog</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Romosozumab</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Unknown to me</td>
<td>6 (9)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>BP: bisphosphonate.  
<sup>b</sup>IV: intravenous.  
<sup>c</sup>PTH: parathyroid hormone.

### eHLQ, SUS, and SUTAQ

A total of 90 participants completed the eHLQ section of the Q0 questionnaire. Mean (SD) scores reported ranged from 2.6 (0.5) to 3.1 (0.6) across the 7 domains (Table 4). A definition of each domain can be found in Multimedia Appendix 1.

A total of 79 participants responded to the SUS questionnaire with a mean (SD) score of 74.7 (14.4), while 74 participants responded to the SUTAQ, with mean (SD) scores for domains 1, 2, and 6 being 3.4 (1.2), 4.5 (1.1), and 4.1 (1.2), respectively (Table 5). Answers to individual questions of the SUTAQ are presented in Table 6.

The covariate “occupation” had an effect on eHLQ domain 4 ($P=.04$) with an adjusted $R^2$ of 0.036 (Tables 7 and 8). No effects of the other covariates were found for the remaining domains of the eHLQ. No correlation was found between the SUS score and age ($P=.37$).

Table 4. eHealth Literacy Questionnaire (eHLQ, N=90).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Value, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eHLQ1: using technology to process health information</td>
<td>2.9 (0.6)</td>
</tr>
<tr>
<td>eHLQ2: understanding of health concepts and language</td>
<td>3.1 (0.5)</td>
</tr>
<tr>
<td>eHLQ3: ability to actively engage with digital services</td>
<td>3.0 (0.6)</td>
</tr>
<tr>
<td>eHLQ4: feel safe and in control</td>
<td>3.1 (0.5)</td>
</tr>
<tr>
<td>eHLQ5: motivated to engage with digital services</td>
<td>2.9 (0.6)</td>
</tr>
<tr>
<td>eHLQ6: access to digital services that work</td>
<td>2.9 (0.5)</td>
</tr>
<tr>
<td>eHLQ7: digital services that suit individual needs</td>
<td>2.6 (0.6)</td>
</tr>
</tbody>
</table>

Table 5. System Usability Score (SUS, n=80) and Service User Technology Acceptability Questionnaire (SUTAQ, n=73) scores at 2 months (Q2).

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Value, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS</td>
<td>74.6 (15.3)</td>
</tr>
<tr>
<td>SUTAQ (scores from 1 to 6)</td>
<td></td>
</tr>
<tr>
<td>Domain 1</td>
<td>4.1 (1.2)</td>
</tr>
<tr>
<td>Domain 2</td>
<td>3.4 (1.1)</td>
</tr>
<tr>
<td>Domain 3</td>
<td>4.5 (1.1)</td>
</tr>
</tbody>
</table>
Table 6. Service User Technology Acceptability Questionnaire individual statements (n=73).

<table>
<thead>
<tr>
<th>Domain 1</th>
<th>Agreement with statement, n (%)</th>
<th>1a</th>
<th>2a</th>
<th>3a</th>
<th>4a</th>
<th>5a</th>
<th>6a</th>
</tr>
</thead>
<tbody>
<tr>
<td>The kit has allowed me to be less concerned about my health status.</td>
<td>12 (16)</td>
<td>3 (4)</td>
<td>19 (26)</td>
<td>13 (18)</td>
<td>16 (22)</td>
<td>10 (14)</td>
<td></td>
</tr>
<tr>
<td>The kit has made me more actively involved in my health.</td>
<td>5 (7)</td>
<td>2 (3)</td>
<td>15 (21)</td>
<td>21 (29)</td>
<td>16 (22)</td>
<td>14 (19)</td>
<td></td>
</tr>
<tr>
<td>The kit allows the people looking after me, to better monitor me and my condition.</td>
<td>15 (21)</td>
<td>14 (19)</td>
<td>13 (18)</td>
<td>11 (15)</td>
<td>13 (18)</td>
<td>7 (10)</td>
<td></td>
</tr>
<tr>
<td>The kit can be or should be recommended to people in a similar condition to mine.</td>
<td>2 (3)</td>
<td>1 (1)</td>
<td>6 (8)</td>
<td>7 (10)</td>
<td>24 (33)</td>
<td>33 (45)</td>
<td></td>
</tr>
<tr>
<td>The kit can certainly be a good addition to my regular health or social care.</td>
<td>4 (5)</td>
<td>1 (1)</td>
<td>7 (10)</td>
<td>17 (23)</td>
<td>20 (27)</td>
<td>24 (33)</td>
<td></td>
</tr>
</tbody>
</table>

Domain 2

<table>
<thead>
<tr>
<th>Agreement with statement, n (%)</th>
<th>1a</th>
<th>2a</th>
<th>3a</th>
<th>4a</th>
<th>5a</th>
<th>6a</th>
</tr>
</thead>
<tbody>
<tr>
<td>The kit I received has saved me time in that I did not have to visit my GP clinic or other health or social care professional as often.</td>
<td>12 (16)</td>
<td>6 (8)</td>
<td>14 (19)</td>
<td>18 (25)</td>
<td>15 (21)</td>
<td>8 (11)</td>
</tr>
<tr>
<td>The kit I received has increased my access to care (health or social care professionals)</td>
<td>12 (16)</td>
<td>7 (10)</td>
<td>19 (26)</td>
<td>22 (30)</td>
<td>10 (14)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>The kit I received has helped me to improve my health</td>
<td>6 (8)</td>
<td>7 (10)</td>
<td>13 (18)</td>
<td>25 (34)</td>
<td>16 (22)</td>
<td>6 (8)</td>
</tr>
<tr>
<td>The kit has made it easier to get in touch with health and social care professionals.</td>
<td>15 (21)</td>
<td>8 (11)</td>
<td>20 (27)</td>
<td>16 (22)</td>
<td>12 (16)</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

Domain 6

<table>
<thead>
<tr>
<th>Agreement with statement, n (%)</th>
<th>1a</th>
<th>2a</th>
<th>3a</th>
<th>4a</th>
<th>5a</th>
<th>6a</th>
</tr>
</thead>
<tbody>
<tr>
<td>The kit has been explained to me sufficiently.</td>
<td>5 (7)</td>
<td>10 (14)</td>
<td>12 (16)</td>
<td>12 (16)</td>
<td>15 (21)</td>
<td>19 (26)</td>
</tr>
<tr>
<td>The kit can be trusted to work appropriately.</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>5 (7)</td>
<td>11 (15)</td>
<td>32 (44)</td>
<td>23 (32)</td>
</tr>
<tr>
<td>I am satisfied with the kit I received.</td>
<td>2 (3)</td>
<td>3 (4)</td>
<td>12 (16)</td>
<td>7 (10)</td>
<td>28 (38)</td>
<td>21 (29)</td>
</tr>
</tbody>
</table>

Note: 1=strongly disagree; 2=moderately disagree; 3=mildly disagree; 4=mildly agree; 5=moderately agree; 6=strongly agree.

Table 7. Summary of regression analysis showing the relationship between occupation and eHealth Literacy Questionnaire domain 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>SE of the estimate</th>
<th>Adjusted $R^2$</th>
<th>$R^2$</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.504</td>
<td>0.036</td>
<td>0.047</td>
<td>0.216$^a$</td>
</tr>
</tbody>
</table>

Note: $^a$Predictors: (Constant). Occupation.

Table 8. ANOVA$^b$ for occupation’s effect on eHealth Literacy Questionnaire domain 4.

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Sum of squares</th>
<th>$df$</th>
<th>Mean square</th>
<th>$F$ ($df$)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1.099</td>
<td>1</td>
<td>1.099</td>
<td>4.315 (1.88)</td>
<td>.04$^b$</td>
</tr>
<tr>
<td>Residual</td>
<td>22.405</td>
<td>88</td>
<td>0.255</td>
<td>_c</td>
<td>_c</td>
</tr>
<tr>
<td>Total</td>
<td>23.504</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $^a$Dependent variable: eHealth Literacy Questionnaire domain 4. 
$^b$Predictors: (Constant). Occupation. 
$^c$Not available.

Discussion

Principal Results

Our analysis revealed eHLQ scores ranging between 2.6 and 3.1 across the 7 domains, with domain 7 “Digital services that suit individual needs” being the lowest-scoring domain and domains 2 “Understanding of health concepts and language” and 4 “Feel safe and in control” being the highest-scoring domains. These findings are comparable to those of another Danish study by Holt et al [20] on eHealth literacy conducted on 246 patients diagnosed with diabetes, other endocrine conditions, or gastrointestinal diseases. Holt et al [20] demonstrated eHLQ scores ranging between 2.6 and 3.1, with
slightly lower scores across the 7 domains. This suggests that our findings are representative of Danish patients with chronic diseases. A Spanish study investigating electronic health literacy in 166 primary care patients, with a median age of 65 (52-78) years, revealed somewhat lower eHLQ scores ranging from 1.7 to 2.8 across the 7 domains [21]. An Australian study of 525 patients at 3 primary care clinics, with a mean age of 56.7, showed eHLQ scores ranging from 2.4 to 3.0 [22]. In comparison with participants from other countries, Danish patients seem to score higher on the eHLQ. These elevated scores among Danish individuals could be attributed to the extensive digitization of public services in Denmark. According to the latest report on digitalization in Denmark (2022), 95% of citizens reported receiving messages from public services through the digital mailbox known as “e-Boks.” Additionally, 74% of citizens aged 15 to 89 years use digital public services at least once a week, and 66% of citizens aged 16 to 74 years have booked a doctor’s appointment via the web and accessed health information through online sources [23]. On examining the eHLQ scores in this study, high scores are observed in domains 1-6, suggesting that Danish patients with osteoporosis are both motivated and able to use digital services related to health. The lowest-scoring domain, domain 7, indicates the current state of technology regarding digital services and points to a potential for developing more individualized services designed to meet the needs of patients.

The statistical analysis revealed a significant but relatively small effect of the covariate “occupation” on the scores of eHealth literacy domain 4 “Feel safe and in control.” This suggests that among the study sample, individuals who are currently working tend to feel safer and more in control when using electronic health services. However, given the small effect size within a limited sample, no definitive conclusions can be drawn. In the field of eHealth literacy, conflicting results have been reported regarding determinants of eHealth literacy scores. A recent systematic review demonstrated an association between eHealth literacy and age, sex, educational level, and family income [24]. Another study conducted by Arcury et al [25] failed to find a connection between eHealth literacy and sociodemographic factors. However, they did discover associations between eHealth literacy and the number of e-devices owned, as well as computer stress [25]. A related finding was reported by Richter et al [26], who concluded that more time spent on the internet was associated with higher eHealth literacy among 453 patients with moderate to high cardiovascular risk. These findings suggest that factors other than sociodemographic variables play a role in eHealth literacy.

The app “My Bones” aims to assist patients in managing osteoporosis. To achieve this objective, it is crucial for the app to be both usable and well received by its users, prompting an investigation into its usability and acceptability. The analysis of the SUS questionnaire at the 2-month mark resulted in a mean (SD) score of 70 indicates a user-friendliness rating of approximately 70. The statistical analysis showed no correlation with age, contradicting the finding by Bangor et al that indicated a slightly negative correlation between the SUS score and age [27]. This discrepancy could be attributed to both the small sample size and the higher average age of this study population.

Analyzing the data from the 3 SUTAQ domains revealed a mean (SD) score of 4.1 (1.2) for domain 1 “Enhanced care,” which indicates a general agreement that the app “My Bones” can improve the care patients receive. The mean (SD) score of 4.5 (1.1) for domain 6 “Satisfaction” suggests acceptance and satisfaction with the app. However, the mean (SD) score of 3.4 (1.1) for domain 2 “Increased accessibility” indicates a lower level of agreement with beliefs that the app has facilitated the receipt of care from health care professionals.

In a Danish study involving 68 patients recruited from primary care and an outpatient clinic, the acceptability of a telehealth service was assessed using the SUTAQ. The scores obtained for domains 1, 2, and 6 were 5.0, 4.2, and 5.5, respectively [28]. Another Danish study focused on telehealth services for patients with chronic obstructive pulmonary disease, diabetes, and inflammatory bowel disease, and pregnant women with either diabetes or a need for enhanced care, demonstrated similar results. Scores for domain 1 ranged from 4.4 to 5.1, domain 2 from 3.8 to 4.7, and domain 6 from 5.2 to 5.6 [29]. Both studies exhibited higher scores across all 3 domains, indicating greater satisfaction and acceptance with the telehealth services compared with the “My Bones” app. One possible explanation for the lower scores in this study is that the SUTAQ was originally designed to assess the acceptability of telehealth services involving remote monitoring by health care professionals [19]. However, the app evaluated in our study does not function as a telehealth system, as it lacks monitoring, data collection by patients, and direct interaction with care personnel. Therefore, the applicability of SUTAQ results in describing the acceptability of the app among our study population may be limited. Certain statements within SUTAQ domains 1 and 2 are rendered irrelevant by the nature of the app. For instance, one statement in domain 2 referred to increased access to care through the received kit, which is not applicable because the app does not connect users with health care professionals. Similar issues were observed in other statements within these domains.

Nevertheless, it is worth noting that specific statements regarding the app’s usefulness showed promising results. Among the 74 participants who answered the SUTAQ questions, 56% agreed that the app saved them a visit to the general practitioner clinic. In total 65% agreed that the app helped them improve their health, 54% agreed that the app made them less concerned about their health status, and 70% agreed that the app made them more involved in their own health. Furthermore, 84% of participants agreed that the app can be a valuable addition to regular health or social care.

In summary, our findings from both the SUS and SUTAQ questionnaires underscore a high level of satisfaction with the “My Bones” app. While the usefulness of SUTAQ scores may
be questionable, the responses to individual statements indicate that the app holds value for patients with osteoporosis.

**Other Findings**

We discovered that over half of the 95 patients with osteoporosis had a significant gap in their knowledge about their condition, as they were not aware of their T scores. The T score reflects the severity of bone loss and risk of fractures. Knowing the T score can empower patients to take preventive measures in their daily lives to avoid injuries. Osteoporosis is a silent disease that often exhibits no symptoms; thus, the T score becomes a crucial indicator for patients to manage their condition. The fact that half of the patients lack awareness of the T score suggests a communication issue within the Danish health care system. Lack of disease-specific knowledge among patients with osteoporosis is a global issue [3], which might be addressed by apps like “My Bones.” However, because we did not inquire about the T score in our follow-up questionnaires, further research is needed to confirm whether the app can improve T score knowledge among patients with osteoporosis.

**Perspectives**

Based on our findings, it appears that the app “My Bones” has the potential to effectively support the self-management of osteoporosis in this population as a supplement to current health care services. Further investigation should be undertaken to fully assess the app’s ability to enhance self-management among patients.

**Limitations**

This study is subject to several limitations. First, the small population size restricts the generalizability of our findings. Additionally, individuals who are already accustomed to using smartphones and health management apps on a daily basis may have been more likely to participate, potentially introducing a bias toward higher eHealth literacy, acceptability, and satisfaction with the app. Recruitment primarily occurred at a single outpatient clinic and through online advertisements on the patient organization’s homepage and a Facebook page, introducing a sampling bias as most patients with osteoporosis in Denmark are typically treated at general practitioner clinics.

The use of questionnaires also presents limitations. Self-reported data on fractures, T scores, and medication are susceptible to response bias and may be less reliable than data obtained from registries. Conversely, the questions on T score and medication were specifically designed to provide insight into the patients’ understanding of their disease.

Despite these limitations, our study provides valuable insights into the eHealth literacy of patients with osteoporosis and the acceptability and usability of the first publicly available osteoporosis management app in Denmark.

**Conclusions**

We uncovered a high level of eHealth literacy, indicating that Danish patients with osteoporosis possess both the motivation and ability to use mHealth services. The app demonstrated acceptable usability and garnered general satisfaction among users. These findings bolster the viability of an app for self-management of osteoporosis to support Danish patients. The app holds the potential to reduce visits to the general practitioner clinic, enhance health outcomes, and serve as a valuable addition to regular health or social care. However, further investigation is necessary to thoroughly evaluate its effectiveness in improving self-management of osteoporosis.

**Acknowledgments**

The authors would like to thank Swinburne University for granting permission to use the eHealth Literacy Questionnaire (eHLQ) and Lars Kayser (National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark) for providing assistance in accessing the questionnaire and offering guidance during the research period. Thanks are also extended to Kristian Kidholm (Department of Clinical Research, University of Southern Denmark) and Professor Stanton Newman (University of London) for granting access to the original Service User Technology Acceptability Questionnaire (SUTAQ), and to Julie Hvildt (Health Innovation Centre of Southern Denmark) for providing access to the Danish version of System Usability Scale (SUS). Additionally, the authors would like to express their thanks to OSAIA Health for granting access to the app “My Bones,” owned and operated by OSAIA Health. This research was supported by Union Chimique Belge Nordic (20141114).

**Conflicts of Interest**

MGB has received speaker honorariums from Union Chimique Belge (UCB), Gedeon Richter, and Novo Nordisk. MFH has received grants from Amgen, UCB, and the Ellab-Foundation; received speaker honorariums from UCB and Amgen; and participated in advisory boards with UCB and Novo Nordic. BMTS and MR declare that they have no conflict of interest.

Multimedia Appendix 1

Scale names and construct definitions of the eHealth Literacy Questionnaire (eHLQ) and visuals of the app “My Bones”:

[DOCX File, 743 KB - ijmr_v13i1e53995_app1.docx ]

**References**


Abbreviations

DXA: dual-energy x-ray absorptiometry
dHLF: eHealth Literacy Framework
eHLQ: eHealth Literacy Questionnaire
IPAQ: International Physical Activity Questionnaire
mHealth: mobile health
SUS: System Usability Scale
SUTAQ: Service User Technology Acceptability Questionnaire

©Magnus Grønlund Bendtsen, Bodil Marie Thuesen Schönwandt, Mette Rubæk, Mette Friberg Hitz. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 01.04.2024. 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.
Intramural Health Care Through Video Consultations and the Need for Referrals and Hospital Admissions: Retrospective Quantitative Subanalysis of an Evaluation Study

Katharina Schmalstieg-Bahr1,2, MD; Miriam Giovanna Colombo3, MPH; Roland Koch3, MD; Joachim Szecsenyi4, MD; Friedrich Völker2; Eva Elisabeth Blozik2,5, MD; Martin Scherer1,2, MD

1Department of General Practice and Primary Care, University Medical Center Eppendorf, Hamburg, Germany
2A+ Videoclinic GmbH, Gräfelfing, Germany
3Institute for General Practice and Interprofessional Care, University Hospital Tübingen, Tübingen, Germany
4Department of General Practice and Health Services Research, University Hospital Heidelberg, Heidelberg, Germany
5Institute of Primary Care, University Hospital Zurich, Zurich, Switzerland

Corresponding Author:
Katharina Schmalstieg-Bahr, MD
Department of General Practice and Primary Care
University Medical Center Eppendorf
Martinistrasse 52
Bldg. W37, 5th Fl
Hamburg, 2046
Germany
Phone: 49 40 7410 52400
Fax: 49 40 7410 4022
Email: k.schmalstieg-bahr@uke.de

Abstract

Background: In comparison to the general population, prison inmates are at a higher risk for drug abuse and psychiatric, as well as infectious, diseases. Although intramural health care has to be equivalent to extramural services, prison inmates have less access to primary and secondary care. Furthermore, not every prison is constantly staffed with a physician. Since transportation to the nearest extramural medical facility is often resource-intensive, video consultations may offer cost-effective health care for prison inmates.

Objective: This study aims to quantify the need for referrals to secondary care services and hospital admissions when video consultations with family physicians and psychiatrists are offered in prison.

Methods: In 5 German prisons, a mixed methods evaluation study was conducted to assess feasibility, acceptance, and reasons for conducting video consultations with family physicians and psychiatrists. This analysis uses quantitative data from these consultations (June 2018 to February 2019) in addition to data from a sixth prison added in January 2019 focusing on referral and admission rates, as well as reasons for encounters.

Results: At the initiation of the project, 2499 prisoners were detained in the 6 prisons. A total of 435 video consultations were conducted by 12 physicians (3 female and 7 male family physicians, and 2 male psychiatrists during the study period). The majority were scheduled consultations (341/435, 78%). In 68% (n=294) of all encounters, the patient was asked to consult a physician again if symptoms persisted or got worse. In 26% (n=115), a follow-up appointment with either the video consultant or prison physician was scheduled. A referral to other specialties, most often psychiatry, was necessary in 4% (n=17) of the cases. Only in 2% (n=8) of the consultations, a hospital admission was needed. Usually, hospital admissions were the result of unscheduled consultations, and the videoconferencing system was the method of communication in 88% (n=7) of these cases, while 12% (n=1) were carried out over the phone. Reasons for admissions were severe abdominal pain, hypotension, unstable angina or suspected myocardial infarction, or a suspected schizophrenic episode.

Conclusions: Most scheduled and unscheduled consultations did not require subsequent patient transport to external health care providers. Using telemedicine services allowed a prompt patient-physician encounter with the possibility to refer patients to other specialties or to admit them to a hospital if necessary.
in 2022 approximately 56,000 people were incarcerated in a total of 172 German prisons. While 19,339 (34%) out of 56,557 inmates face imprisonment of over 1 year to 4 years, 5340 (9%) out of 56,557 inmates serve prison sentences of 5 years or longer [1,2]. By law, these prisoners have the same right to access health care as patients with statutory health insurance outside the correctional system. However, caring for incarcerated patients is challenging. They are more likely to experience alcohol or drug abuse, mental illnesses, or communicable diseases, such as hepatitis C or HIV infections [3,4], and not every prison is constantly staffed with a physician on-site. Specialized (secondary care) services are less available to inmates since a resource-intensive transport to the next extramural facility has to be organized. Telemedicine offers the potential to close this gap and improve intramural health care [5]. In this context, the collective term telemedicine describes heterogeneous concepts that aim at providing medical diagnostics, therapy, and rehabilitation despite physical distance or time lag [6]. In many countries, it has been used to facilitate or enhance intramural care [7-10], but it has not been implemented on a broad scale in the German correctional system yet.

This study aims to assess the need and the reasons for referrals to secondary care services, as well as hospital admissions, when scheduled or unscheduled telemedicine consultations with family physicians and psychiatrists are offered in prison.

Methods

Overview

The pilot project to establish video consultations in German prisons was initiated by the Ministry of Justice Baden-Württemberg in cooperation with A+ Videoclinic (VC), a provider of telemedical services, and initially, involved 5 German prisons. Between June and December 2018, the pilot project was evaluated in terms of feasibility and acceptance of the video consultations, as well as consultation reasons by conducting a mixed methods evaluation study. Quantitative and qualitative data were collected through site visits in the prisons, questionnaires, semistructured interviews, and consultation documentation. Further details are reported elsewhere [11]. This analysis is a retrospective subanalysis using a quantitative VC data set that was generated during the evaluation study period containing the information depicted in Table 1, as well as additional data from a sixth prison, which was not part of the initial pilot project.

All 6 prisons were located in the federal state of Baden-Württemberg, Germany. Inmates were male and female (adults and adolescents) and 18 years of age and older. Participation in the pilot study was voluntary. Patients could choose either a video consultation or regular medical care. If the patient opted for a video consultation, he or she had to sign an informed consent form.

Consultations that were conducted between June 2018 and February 2019 were analyzed in this study. The videoconferencing system (VCS) was the preferred method of communication. The phone was used in case of any technical problems with the VCS. Consultations were carried out by a team of 12 physicians (3 female and 7 male family physicians, and 2 male psychiatrists) employed by the telemedicine provider. Scheduled encounters were conducted during fixed weekly timeslots—either with VC-family physicians or a VC psychiatrist. Outside of these consultation hours, prison nursing staff could reach the on-call VC-family physician 24 hours 7 days per week. These patient-physician-contacts outside of consultation hours were counted as unscheduled consultations. Depending on the time of contact, there was not always a trained nurse present in prison. If the on-call family physician required help regarding a psychiatric problem, he or she could contact a VC psychiatrist.

The VC provided the telemedical infrastructure for VC physicians and prisons. Physicians documented the consultations electronically with a VC laptop using a virtual private network to access the VC software called Videoclinic Portal (Videoclinic) developed by the Videoclinic. The participating prisons were also equipped with VC laptops to receive the documentation. No remote medical devices, such as stethoscopes, that would have allowed the physician to directly auscultate a patient were used during the pilot study. Either prison nursing staff or correctional officers were present during each encounter. Prison nursing staff could obtain the patient’s vital signs (pulse, blood pressure, temperature, and oxygen saturation) and write an electrocardiogram if necessary and available. A total of 4 out of 6 prisons had an electrocardiogram on-site. Further details on the medical equipment available in the 5 prisons that were part of the initial pilot project can be found elsewhere [11].

For this study, an anonymized data set was exported from the VC software containing the variables depicted in Table 1. The data analysis for this study comprised descriptive statistical methods and was performed using Microsoft Excel 2016.
Table 1. Data provided by A+ Videoclinic.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of consultation</td>
<td>Date and time of the encounter</td>
</tr>
<tr>
<td>Prison</td>
<td>Name of the prison</td>
</tr>
<tr>
<td>Physician</td>
<td>Treating physician (pseudonymized number D1-12)</td>
</tr>
<tr>
<td>Medical specialty</td>
<td>Family medicine</td>
</tr>
<tr>
<td></td>
<td>Psychiatry</td>
</tr>
<tr>
<td>Assessment</td>
<td>Current assessment (free text)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Current diagnosis (free text and International Classification of Diseases, 10th revision code)</td>
</tr>
<tr>
<td>Plan</td>
<td>Recommended treatment (free text)</td>
</tr>
<tr>
<td>Medication</td>
<td>Current medication if any prescribed (free text)</td>
</tr>
<tr>
<td>Type of consultation</td>
<td>Scheduled, Unscheduled</td>
</tr>
<tr>
<td>Method of communication</td>
<td>Video, Phone</td>
</tr>
<tr>
<td>Interpreter</td>
<td>Foreign language if interpreter was used</td>
</tr>
<tr>
<td>Follow-up</td>
<td>No further treatment, Follow-up if symptoms persist or worsen, Planned follow-up appointment, Referral, Hospital admission</td>
</tr>
</tbody>
</table>

Ethics Approval

Ethics approval for the evaluation of the pilot project was obtained from the ethics committee of the Eberhard-Karls-University Tübingen (728/2018BO1) and the ethics committee of the State Medical Association Baden-Württemberg (F-2018-054) [11].

Results

Baseline Characteristics of the Participating Prisons

The ratio of prisoners to medical staff differed between the 6 prisons. Table 2 shows the sex and number of inmates of each prison at the initiation of the pilot study, as well as the number of physicians and nurses. Except for P4, prison staff comprised at least 1 physician. In P4, 2 external physicians offered scheduled consultation hours twice a week. Scheduled consultations with external physicians from other specialties (not family medicine) varied greatly. In most prisons, regular visits from a dentist were established. P2 offered appointments with a gynecologist and psychiatrist. In addition to dental care and the care provided by the prison physician, P6 offered consultations with a dermatologist and surgeon. Furthermore, nursing staff from P6 was able to contact an external psychiatrist if needed.

Table 2. Occupancy and staff per prison.

<table>
<thead>
<tr>
<th>Prison</th>
<th>Characteristics</th>
<th>Number of prisoners (n=2499), n (%)</th>
<th>Number of physicians (n=12), n (%)</th>
<th>Number of nurses (n=69), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Male adults</td>
<td>350 (14)</td>
<td>2 (17)</td>
<td>9 (13)</td>
</tr>
<tr>
<td>P2</td>
<td>Female adults or adolescents</td>
<td>350 (14)</td>
<td>1 (8)</td>
<td>6 (9)</td>
</tr>
<tr>
<td>P3</td>
<td>Male adolescents</td>
<td>395 (16)</td>
<td>4 (33)</td>
<td>9 (13)</td>
</tr>
<tr>
<td>P4</td>
<td>Male adults</td>
<td>52 (2)</td>
<td>0 (0)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>P5</td>
<td>Male adults</td>
<td>772 (31)</td>
<td>4 (33)</td>
<td>22 (32)</td>
</tr>
<tr>
<td>P6</td>
<td>Male adults</td>
<td>580 (23)</td>
<td>1 (8)</td>
<td>20 (29)</td>
</tr>
</tbody>
</table>
Referrals Following Video Consultations

From June 2018 to February 2019, VC physicians conducted 435 consultations. Out of that, 78% (n=341) of all consultations were scheduled, and the remainder were unscheduled (94/435, 22%). In 68% (n=294), patients were asked to consult a physician again if symptoms persisted or got worse. In 26% (n=115) a follow-up appointment with the video consultant or prison physician was scheduled. A referral to other specialties was necessary in 4% (n=17). The VCS was the method of communication in all 17 cases during which a referral was necessary. A total of 3 of these encounters were unscheduled. Patient transport to an extramural facility was indicated in none of these cases since the secondary care physician could either be contacted via video consultation, was able to come to the prison, or was employed by the prison but off duty during the initial encounter (Table 3).

Table 3. Overview of the referrals to other specialties (n=17).

<table>
<thead>
<tr>
<th>Type of consultation</th>
<th>Medical specialty (physician number)</th>
<th>Diagnosis</th>
<th>Medical specialty referred to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unscheduled</td>
<td>FMa (D7)</td>
<td>Drug abuse, opiate withdrawal</td>
<td>Psychiatry (coming to site)</td>
</tr>
<tr>
<td>Scheduled</td>
<td>PSYb (D8)</td>
<td>Adjustment disorder</td>
<td>Child and adolescent psychiatry</td>
</tr>
<tr>
<td>Unscheduled</td>
<td>FM (D9)</td>
<td>Acneiform dermatitis</td>
<td>Dermatology (video consultation)</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D5)</td>
<td>Nausea, suspected adverse reaction, suspected melena</td>
<td>Rheumatology</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D10)</td>
<td>(missing data)</td>
<td>(missing data)</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D10)</td>
<td>Presbyopia</td>
<td>Ophthalmology</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D10)</td>
<td>(missing data)</td>
<td>(missing data)</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D5)</td>
<td>Ankle sprain R</td>
<td>Physician licensed to treat work accidents (German: Durchgangsarzt)</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D10)</td>
<td>(missing data)</td>
<td>(missing data)</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D5)</td>
<td>Adverse drug reaction</td>
<td>Urology</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D10)</td>
<td>Toothache</td>
<td>Dentistry</td>
</tr>
<tr>
<td>Unscheduled</td>
<td>FM (D9)</td>
<td>Prison admission exam</td>
<td>Psychiatry or drug counseling (prison psychiatrist)</td>
</tr>
<tr>
<td>Scheduled</td>
<td>PSY (D8)</td>
<td>Chronic arm pain after car accident, insomniaa</td>
<td>Neurology</td>
</tr>
<tr>
<td>Scheduled</td>
<td>PSY (D8)</td>
<td>Liver cirrhosis, suspected ascitesb</td>
<td>Gastroenterology</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D5)</td>
<td>Priapism</td>
<td>Urology</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D5)</td>
<td>First degree hemorrhoids</td>
<td>Proctologyd</td>
</tr>
<tr>
<td>Scheduled</td>
<td>FM (D5)</td>
<td>First degree hemorrhoids</td>
<td>Proctologyd</td>
</tr>
</tbody>
</table>

aFM: family medicine.
bPSY: psychiatry.
cNo diagnosis coded—information taken from medical history.
dReferral only necessary if symptoms persist despite ordered treatment.

Hospital Admissions Following Video Consultations

In 2% (n=8) of the cases, a hospital admission was required. These cases were independent of the 17 cases that required a referral to another specialty in an ambulatory care setting. Hospital admission was usually the result of an unscheduled consultation (7/8, 88%), and the VCS was used in 88% (n=7). Gastrointestinal problems or pain were the most common reason for admission (4/8, 50%) and 6 VC physicians (5 family physicians and 1 psychiatrist) conducted the encounters (Table 4). An interpreter was not needed in any of these consultations.
Table 4. Overview of hospital admissions (n=8).

<table>
<thead>
<tr>
<th>Type of consultation</th>
<th>Method of communication</th>
<th>Medical specialty (physician number)</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unscheduled</td>
<td>Video</td>
<td>FM (D1)</td>
<td>Suspected tuberculosis, hypotension</td>
</tr>
<tr>
<td>Unscheduled</td>
<td>Video</td>
<td>FM (D7)</td>
<td>Unstable angina</td>
</tr>
<tr>
<td>Unscheduled</td>
<td>Video</td>
<td>FM (D1)</td>
<td>Severe hypotension suspected due to antipsychotics</td>
</tr>
<tr>
<td>Unscheduled</td>
<td>Video</td>
<td>FM (D11)</td>
<td>Gastrointestinal hemorrhage</td>
</tr>
<tr>
<td>Unscheduled</td>
<td>Phone</td>
<td>FM (D2)</td>
<td>Abdominal pain, kidney stones</td>
</tr>
<tr>
<td>Unscheduled</td>
<td>Video</td>
<td>FM (D2)</td>
<td>Gallbladder disease</td>
</tr>
<tr>
<td>Unscheduled</td>
<td>Video</td>
<td>FM (D4)</td>
<td>Abdominal pain (diagnostic: myocardial infarction)</td>
</tr>
<tr>
<td>Scheduled</td>
<td>Video</td>
<td>PSY (D8)</td>
<td>Suspected schizophrenia</td>
</tr>
</tbody>
</table>

a FM: family medicine.
b PSY: psychiatry.

Hospital Admission and Referrals per Prison and Physician

Almost half of the video consultations (194/435, 45%) were conducted in prison P2 (Table 5). It was the only prison with female inmates (adults and adolescents). At the time of the initial visit to the participating prisons in the framework of the pilot project (September 2018), the number of inmates was similar to P1 (male adults) and P3 (male adolescents; Table 2).

A referral following a video consultation was needed in all prisons apart from P1 and P6. P3 had the highest number of referrals, but less than 10% of video consultations were conducted there. A hospital admission following a video consultation was required in all prisons except P3 and P4.

VC physicians D1 and D8 conducted more than 60% of all encounters (Table 6). However, referrals and admissions were rather scattered among physicians—except for D5, who initiated 6 (35%) of the referrals.

Table 5. Video consultations, referrals, and hospital admissions per prison.

<table>
<thead>
<tr>
<th>Prison</th>
<th>Video consultations per prison, (n=435), n (%)</th>
<th>Referrals per prison, (n=17), n (%)</th>
<th>Hospital admissions per prison, (n=8), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>41 (9)</td>
<td>0 (0)</td>
<td>2 (25)</td>
</tr>
<tr>
<td>P2</td>
<td>194 (45)</td>
<td>5 (29)</td>
<td>1 (13)</td>
</tr>
<tr>
<td>P3</td>
<td>41 (9)</td>
<td>6 (35)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>P4</td>
<td>75 (17)</td>
<td>4 (24)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>P5</td>
<td>68 (16)</td>
<td>2 (12)</td>
<td>4 (50)</td>
</tr>
<tr>
<td>P6</td>
<td>16 (4)</td>
<td>0 (0)</td>
<td>1 (13)</td>
</tr>
</tbody>
</table>

Table 6. Video consultations, referrals, and hospital admissions per physician.

<table>
<thead>
<tr>
<th>Physician</th>
<th>Video consultations per physician, (n=435), n (%)</th>
<th>Referrals per physician, (n=17), n (%)</th>
<th>Hospital admissions, (n=8), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>143 (33)</td>
<td>1 (6)</td>
<td>2 (25)</td>
</tr>
<tr>
<td>D2</td>
<td>9 (2)</td>
<td>0 (0)</td>
<td>2 (25)</td>
</tr>
<tr>
<td>D3</td>
<td>5 (2)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>D4</td>
<td>18 (4)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>D5</td>
<td>14 (3)</td>
<td>6 (35)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>D6</td>
<td>8 (2)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>D7</td>
<td>2 (1)</td>
<td>1 (6)</td>
<td>1 (13)</td>
</tr>
<tr>
<td>D8</td>
<td>124 (29)</td>
<td>3 (17)</td>
<td>1 (13)</td>
</tr>
<tr>
<td>D9</td>
<td>9 (2)</td>
<td>2 (12)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>D10</td>
<td>97 (22)</td>
<td>4 (24)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>D11</td>
<td>3 (1)</td>
<td>0 (0)</td>
<td>1 (13)</td>
</tr>
<tr>
<td>D12</td>
<td>3 (1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
**Discussion**

**Principal Findings**
Most scheduled and unscheduled video consultations did not require a subsequent patient transport to an extramural health care provider or facility. A referral was only needed in 4% (n=17) of the cases and hospital admission was only required in 2% (n=2) of 435 cases.

**Comparison to Prior Work**
To our knowledge, this is the first study that focused on referral and hospital admission rates in intramural health care when video consultations with family physicians and psychiatrists are offered. Another study evaluated telemedicine consultations with an emergency room (ER) physician and found that 36% of the patients from a correctional facility required transport to the ER after a video encounter [12]. This rate was higher than the referral and hospital admission rate of this study. If prison physicians were given the possibility to refer patients to a telemedicine satellite facility for subspecialty consults, outpatient visits would increase by 40% in the 2 years after implementation. In contrast, ER visits decreased. The authors interpreted the effect as better access to care [13]. The same authors examined telemedicine programs in the juvenile justice system by measuring, for example, outpatient costs, ER costs, and transportation costs [14], which are parameters that are associated with referrals and admissions. Other studies focused on the treatment of single diseases, such as hepatitis C or diabetes, when telemedicine consultations were offered in prison [15,16].

Other studies have shown that outside of the correctional system home telemonitoring could reduce hospital admission rates in people with chronic obstructive pulmonary disease [17,18] and congestive heart failure [19,20]. Telemedicine consultations with an emergency physician led to reduced transfers from skilled nursing facilities to ERs and subsequently to lower hospital admission rates [21], and Rosner et al [22] showed that using telemedicine reduced readmissions after hip and knee arthroplasties. In contrast, a meta-analysis that evaluated different remote management strategies for patients with inflammatory bowel disease demonstrated a reduction in physician visits but no significant effect on relapse or hospital admission rates [23].

Regarding a reduction of referral rates, previous studies outside of the correctional system generally focused on electronic solutions for primary care providers to contact other specialists and have demonstrated mixed results. Liddy et al [24] reported that implementing an electronic consultation service in Canada that allowed family physicians to communicate with secondary care providers regarding a patient’s care reduced referral rates between 36% and 53%, which was in line with prior findings of that group [25]. However, their randomized controlled trial with 2 study arms (primary care physicians with and without access to the electronic consultation service), showed a significant referral reduction in both arms [26]. Furthermore, according to an online survey among primary care physicians, a phone consultation with an HIV specialist reduced the perceived need to refer the patient to a secondary care provider, although the authors acknowledged that actual referral rates had not been studied [27]. In another study, web-based consultation between primary care physicians and nephrologists did not affect referral rates [28].

In comparison to referral rates of (extramural) family medicine practices without the use of video consultations, a referral rate of 3.8% found in this study is at the lower end of the expected spectrum. Generally, referral rates to secondary care providers vary between practices. A recent study showed an average monthly rate of 20.3% with a range of 0.4%-67.1%. Outside of the correctional system, mental health services were the 10th most common specialty for referrals [29]. Older studies showed mean rates of 1.4% to 37% [30-33]. Variance can also be found regarding hospital admission rates of (extramural) family medicine practices. Mean rates of approximately 50-53 admissions per 1000 patients per year have been reported [34,35], which can only indirectly be compared with the data of this analysis (8 admissions per 435 encounters within 9 months).

International studies showed that telemedicine was able to deliver high-quality and timely primary care for adult and adolescent prison inmates [13,36], reduce costs [37], and facilitate mental health services [38,39]. Using telemedicine to improve access, cost, and quality of secondary care, for example, in the fields of ophthalmology [40,41], cardiology [42], and dentistry [10,43] has also been described before and is still under evaluation [44]. Especially, countries with remote regions outside of metropolitan areas, such as the United States or Australia, have reported the use of telemedicine in their correctional systems [45], but also more densely populated countries deemed the use as beneficial [10,46].

**Strengths and Limitations**
The pilot project was the first broad implementation of telemedicine in the German correctional system. But there are some limitations: first, a possible selection bias has to be considered: prison staff may have chosen to directly call an ambulance or organize a transport to an extramural facility instead of using video consultations for more severe medical cases. If only patients with less severe diseases were seen by VC physicians, the likelihood of required hospital admission was consequently decreased. Similarly, prison staff may have directly scheduled an appointment with a secondary care provider visiting the prison, which also might have reduced the likelihood of referrals. However, it was crucial that not only patients were given a choice to talk to a VC physician or receive usual care, but that prison staff was also free to choose whether or not to contact the VC—especially since 2 authors were VC founding members. Neither prison staff nor patients received incentives for contacting the VC. Second, there was no control group in this study, and therefore, referral and hospital admission rates cannot be compared with regular intramural care. Despite the low rates when using video consultations, a referral or an admission might not have been necessary if an in-person consultation had been done at that time. Third, only data from the VC portal was considered for this analysis. Fourth, it is unknown how many patients refused a video consultation. Fifth, at the beginning of the data collection, some VC physicians did
not complete the entire documentation template, therefore, some data were missing (the type of follow-up was not specified in 1 case, and in 2 cases, no diagnosis or reason for the referral was listed). The documentation improved throughout the project as physicians became more familiar with or received more training regarding the use of the VC portal. Finally, all consultations were carried out by a rather small group of physicians, and therefore, the influence of individual experience and working style cannot be excluded.

**Implication for Practice and Research**
The results show that a referral or admission was only required in a few cases after video consultations were offered in prison. Compared with extramural family medicine practices, the referral rates found in this study were at the lower end of the expected spectrum and support that the implementation of telemedicine in intramural systems on a larger scale should not be postponed or revised due to concerns of high referral and admission rates. Further research, including controlled studies, is needed to explore whether institutional factors contribute to the effectiveness and safety, such as training of staff, use of remote medical devices, and acceptance of telemedicine by inmates, prison (nursing) staff, and physicians working on-site—especially in the light of the fact that the data were generated prior to the COVID-19 pandemic and video consultations became much more common since then.

**Acknowledgments**
The authors thank all participating patients, prison staff, and VC physicians. No machine learning, artificial intelligence, language models, or similar technology has been used for the project or to draft the manuscript. The initial project was funded by the Ministry of Justice Baden-Württemberg. No funding was received for this analysis. The authors acknowledge the financial support from the Open Access Publication Fund of the University Medical Center Hamburg-Eppendorf.

**Data Availability**
The data sets generated or analyzed during this study are not publicly available due to privacy protection regulations but are available from the corresponding author on reasonable request.

**Authors’ Contributions**
MS planned and acquired the project between the Ministry of Justice Baden-Württemberg and the A+ Videoclinic. MS and EEB monitored the telemedical processes and the quality management protocols. RK and MGC collected the data for the evaluation study of the pilot project. KS-B analyzed the data of this study. All authors helped to draft the manuscript and approved the final version.

**Conflicts of Interest**
MS is a founding member of the A+ Videoclinic and the head of the Medicine and Quality Committee. FV is the current chief executive officer. EEB is the acting medical head. JS represented the project to the Medical Association of Baden Württemberg as the medical leader. KS-B is a part-time employee of the A+ Videoclinic. RK and MGC declare no conflict of interest.

**References**


**Abbreviations**

- **ER:** emergency room
- **VC:** video clinic
- **VCS:** videoconferencing system

Edited by T Leung; submitted 08.12.22; peer-reviewed by H Leach, R Koppel, F Prazeres; comments to author 07.01.24; revised version received 29.02.24; accepted 01.03.24; published 28.06.24.

Wenjun Gu1, MPH; Jinhua Wang2, BA; Yunqi Zhang3, MSc; Shaolin Liang3,4*, PhD; Zisheng Ai1*, PhD; Jiyu Li5*, MD

1School of Medicine, Tongji University, Shanghai, China
2Shanghai Institute of Medical Innovation Technology Transfer, Shanghai, China
3STI-Zhilian Research Institute for Innovation and Digital Health, Beijing, China
4Institute for Six-sector Economy, Fudan University, Shanghai, China
5Geriatric Oncology Center, Huadong Hospital Affiliated to Fudan University, Shanghai, China
*these authors contributed equally

Corresponding Author:
Jiyu Li, MD
Geriatric Oncology Center
Huadong Hospital Affiliated to Fudan University
221 West Yan'an Road, Jing'an District
Shanghai, 200040
China
Phone: 86 18918231588
Email: lijiyu@fudan.edu.cn

Abstract

Background: The significant impact of digital health emerged prominently during the COVID-19 pandemic. Despite this, there is a paucity of bibliometric analyses focusing on technologies within the field of digital health patents. Patents offer a wealth of insights into technologies, commercial prospects, and competitive landscapes, often undisclosed in other publications. Given the rapid evolution of the digital health industry, safeguarding algorithms, software, and advanced surgical devices through patent systems is imperative. The patent system simultaneously acts as a valuable repository of technological knowledge, accessible to researchers. This accessibility facilitates the enhancement of existing technologies and the advancement of medical equipment, ultimately contributing to public health improvement and meeting public demands.

Objective: The primary objective of this study is to gain a more profound understanding of technology hotspots and development trends within the field of digital health.

Methods: Using a bibliometric analysis methodology, we assessed the global technological output reflected in patents on digital health published between 2017 and 2021. Using Citespace5.1R8 and Excel 2016, we conducted bibliometric visualization and comparative analyses of key metrics, including national contributions, institutional affiliations, inventor profiles, and technology topics.

Results: A total of 15,763 digital health patents were identified as published between 2017 and 2021. The China National Intellectual Property Administration secured the top position with 7253 published patents, whereas Koninklijke Philips emerged as the leading institution with 329 patents. Notably, Assaf Govari emerged as the most prolific inventor. Technology hot spots encompassed categories such as “Medical Equipment and Information Systems,” “Image Analysis,” and “Electrical Diagnosis,” classified by Derwent Manual Code. A patent related to the technique of receiving and transmitting data through microchips garnered the highest citation, attributed to the patentee Covidien LP.

Conclusions: The trajectory of digital health patents has been growing since 2017, primarily propelled by China, the United States, and Japan. Applications in health interventions and enhancements in surgical devices represent the predominant scenarios for digital health technology. Algorithms emerged as the pivotal technologies protected by patents, whereas techniques related to data transfer, storage, and exchange in the digital health domain are anticipated to be focal points in forthcoming basic research.

KEYWORDS
technology trends; digital health; patent; bibliometric analysis; CiteSpace5.1R8

Introduction

Background

Over the past few decades, the intertwining of health care and digital technology has given rise to significant transformations in the production, alignment, and consumption of health care products [1]. This integration has contributed to the achievement of safer and more cost-effective health care outcomes. The conceptual emergence of terms such as “Digital Health,” “Digital Medicine,” and “Digital Therapeutics” is a direct consequence of this symbiotic relationship. The International Digital Therapeutics Alliance (DTA), founded in 2017 in the United States, stands as a nonprofit industry association comprising stakeholders committed to evidence-based therapeutic interventions aimed at preventing, managing, or treating diseases [2]. As per the DTA’s classification, digital therapeutics represents a specific niche within digital health and digital medicine [3]. Notably, a range of digital therapeutics products is currently available for managing diabetes [4,5], treating patients with social anxiety disorder [6] or neurological disorders [7], addressing mental illness [8], and developing digital biomarkers designed to predict treatment response [9]. These products rely on real-world data, and clinical evidence is imperative to substantiate claims regarding risk, efficacy, and intended use [10].

Nielsen and Sahay [11] conducted a critical examination aimed at identifying gaps in the literature, approaching the analysis from an information systems perspective. Their study, based on the analysis of 342 articles published in interdisciplinary digital health research journals, revealed that these studies tended to deviate from the complexities inherent in real-life settings within health care organizations, particularly in relation to the characteristics of digital technologies and the context of their use. Notably, the literature on digital health primarily emphasized the processing power of digital technology and its potential effects [11]. In line with the observations from the study by Nielsen and Sahay [11], a bibliometric analysis conducted by Keng Yang et al [12] in 2022 spanning the period from 1998 to 2021 demonstrated a rapid expansion in the number of publications on digital health. The findings suggested that heightened awareness of digital health had the potential to improve health outcomes, bridge digital gaps, and reduce health disparities [12]. Supporting this perspective, Ahmadvand et al [13] used “digital health” as a keyword for their bibliometric analysis, covering articles published between January 2000 and August 2019 in JMIR Publications journals. Their study revealed a significant increase in the number of articles focusing on “digital health” over a 9-year period, with “mHealth” emerging as the most frequently used keyword within this domain [13]. In addition, Gupta et al [14] conducted a scientometric assessment of digital health research, examining 6981 global publications sourced from the Scopus database during the period 2007 to 2016. Their research underscored medicine as the dominant topic, constituting the largest publication share in digital health research at 54% [14].

Previous studies primarily focused on published papers, neglecting the distinctive value provided by patents as a special and unique source of knowledge, given that a significant amount of data and information contained in patents is not made available through alternative channels. The patent system, designed to stimulate innovation, concurrently ensures that the benefits of inventive efforts are made accessible to the public. In the field of digital health, considering the industry’s rapid and iterative development, the patent system must leverage digital technology to safeguard algorithms, software, and advanced surgical instruments. Apart from furnishing information on digital health patents, the patent system serves an additional role by granting other researchers access to patent information, thereby facilitating enhancements in existing technologies.

Bibliometric analysis serves as a powerful tool for examining research trends, identifying prolific authors, understanding demographics, and exploring related information within specific fields [15]. CiteSpace5.1R8, a widely used scientific mapping software application, is adept at analyzing hot spots and trends, visually illustrating a systematic understanding of the past across various domains [16]. Bibliometrics, along with visual analysis, of patent landscapes in the fields of digital technology within public health is relatively scarce. Therefore, this study predominantly relies on this methodology to investigate granted patents related to digital technology in health care, shedding light on its technical status and prevailing tendencies.

Objectives

The research uses the bibliometric method to delve into the annual volume of patent applications and grants, scrutinizing information related to countries, inventors, patentees, cited patents, and Derwent Manual Code (DMC) classifications of digital technologies. The primary objective of this study is to gain a more profound understanding of technology hot spots and development trends within the field of digital health.

Methods

Data Collection

The data for this study were sourced from the Derwent Innovation Index, a database that amalgamates patent citations from the Derwent Patents Citation Index with additional patent data indexed from >50 patent issuing agencies in the Derwent World Patent Index (1963 to present). Adhering to the procedural guidelines by Chen [17], we retrieved patents associated with “digital health,” “digital medicine,” and “digital therapeutics,” along with a selection of other pertinent topics. Our screening strategies incorporated DMC “T01-J06A,” signifying “medical equipment and information systems;” and “S05-G02G,” representing “hospital equipment with medical IT systems;” based on the categorization of DMC. DMC serves as a simplified classification system designed to categorize patent documents across all technologies. This coding system streamlines the categorization and indexing method used by
Derwent for all covered patents, ensuring efficiency [18,19]. Figure 1 illustrates our specific screening strategies in the flowchart for data collection.

**Figure 1.** The flowchart for data collection.

The patent retrieval process involved several steps to compile a comprehensive data set related to digital health. Initially, we extracted patents with topics relevant to digital health using the topic tag for string retrievals through titles and abstracts of patents in the Derwent Innovation Index. The first step (#1) used the following query: Topics=(“digital therapeutics” OR “medical software” OR “digital medic*” OR “digital health” OR “digital diagnos*” OR “digital surgery” OR “digital biomarker” OR “software driven medical intervention” OR “data driven medical intervention” OR “software medical devices” OR “clinical data security”). Subsequently, we retrieved patents using DMC. The query command for the second step (#2) was Manual Codes=(T01-J06A OR S05-G02G). In the third step, we combined the commands from #1 and #2. Patent publication year=2017 to 2021. The retrieved patents were published from 2017 to 2021, and the duplicate are excluded. n=93,780. The query command for the fourth step (#4) was Patent type=Invention patent. n=43,062. The query command for the fifth step (#5) was Legal status of patents=patent granted and valid. Patents included in patent metrology analysis n=15,763.

OR “digital diagnosis*” OR “digital surgery” OR “digital biomarker” OR “software-driven medical intervention” OR “data-driven medical intervention” OR “software medical devices” OR “clinical data security”). Subsequently, we retrieved patents using DMC. The query command for the second step (#2) was Manual Codes=(T01-J06A OR S05-G02G). In the third step, we combined the commands from
#1 OR #2 and limited the publication date of patents from 2017 to 2021. This timeframe was chosen because significant developments in digital health occurred in 2017, including the establishment of the International DTA, release of guidelines by the Chinese State Food and Drug Administration, publication of the World Health Organization recommendations on digital technologies for Tuberculosis care, and the US Food and Drug Administration approval of the first prescription digital therapy. We retrieved 93,780 published patents using this criterion. Subsequently, in the fourth step, we focused on invention patents, as they generally hold greater value in most countries, resulting in 43,062 invention patents included in the sample. As expired patents lack legal force, our analysis primarily considered patents in force to reflect the latest trends in digital technology in public health. We included patents from the previous step (#4) with a granted and valid legal status, resulting in a final data set of 15,763 patents included in our study sample. According to our search methods, patents from the African region have not been retrieved in this study.

As outlined in the introduction to the Derwent Innovation Index database, the earliest time stamp for data collection extended back to 1963. The initiation of patent retrieval for published patents began in 2017. Given that some patents published in 2017 might have been granted in 2022, the conclusion of the study period was set at the year 2022. The data retrieval process was completed on September 22, 2022.

Analysis Methodologies

In conducting this study, CiteSpace 5.1 R8 SE and Excel 2016 software were used. By configuring parameters such as time slicing and thresholds, diverse results could be obtained [20,21]. Furthermore, CiteSpace offered the capability to visualize the outcomes of the analysis [22], using linkages and nodes to illustrate the quality of various elements and their interrelationships [23]. The software also had the ability to cluster items using statistical techniques such as log-likelihood ratio and assess the burst of themes to identify DMCs with prolonged bursts and high strength [24]. CiteSpace contributes to enhanced clarity and interpretability of visualizations compared to other tools, thereby reducing cognitive load on users as they explore significant trends and turning points in a technological framework [25,26].

The chosen records were exported from the Derwent Innovation Index in plain text format and subsequently imported into CiteSpace for comprehensive analysis and visualization of cited patents, research hot spots, and frontiers. A total of 15,763 patent records, encompassed by the CiteSpace version, were retrieved based on their patent publishing timeframe spanning from 2017 to 2022. The data were segmented into 1-year time slices, and the top 50 most cited strings were extracted from each slice. The thematic analysis incorporated various fields from the plain texts, encompassing inventors, cited patents, patent assignees, cited authors, DMC, and International Patent Classification (IPC). Node types were specified as country, institution, author, and category, and the results were visualized in this specified order. This approach allowed for a detailed exploration of relationships and patterns within the data set, providing insights into the geographical, institutional, and authorship aspects of digital health patents.

The chosen records were exported from the Derwent Innovation Index in a table format and subsequently imported into Excel 2016 for thorough analysis. The study focused on examining the annual volume of applications, granted patents, the most productive countries, institutions, inventors, highly cited patents, and institutes. This approach in Excel allowed for a systematic exploration of quantitative aspects related to digital health patents, providing insights into the dynamics of patent activity over time and identifying key contributors and influential patents in the field.

Results

Applications and Granted Trends Across Diverse Sectors

We conducted an analysis of all 15,763 granted patents related to digital health published between 2017 and 2021. In 2017, a total of 48 patents were granted, followed by a surge to 428 granted patents in 2018. As illustrated in Figure 2, the annual applications for patents in the digital health domain experienced a sharp increase from 2017 to 2019, followed by a slight stagnation in 2020. However, a renewed upward trend was observed from 2020 to 2021. A general upward trajectory was noted in granted patents from 2017 to 2022. The fluctuation in 2022 may be attributed to the timing of data retrieval, as patents published between 2017 and 2021 might not have been granted by September 22, 2022.
Geographical Distribution
Applicants have the option to apply for patents through intellectual property offices (IPOs) in various countries or regions, thereby extending the scope of protection for their technological innovations in the field of digital health. Table 1 and Figure 3 present the top 10 IPOs for published patents and the top 10 patent applicant countries concerning patent publications. These insights offer a comprehensive view of the global landscape in terms of patent activities related to digital health, highlighting the key contributors and regions involved in the innovation and protection of digital health technologies.

Table 1. Top 10 intellectual property offices for published patents (N=15,763).

<table>
<thead>
<tr>
<th>Intellectual property office</th>
<th>Published patents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>China National Intellectual Property Administration</td>
<td>7253 (46.01)</td>
</tr>
<tr>
<td>United States Patent and Trademark Office</td>
<td>4052 (25.71)</td>
</tr>
<tr>
<td>Japanese patent office</td>
<td>2174 (13.79)</td>
</tr>
<tr>
<td>European patent office</td>
<td>1242 (7.88)</td>
</tr>
<tr>
<td>IP Australia</td>
<td>244 (1.55)</td>
</tr>
<tr>
<td>Inspirations Property Institution</td>
<td>228 (1.45)</td>
</tr>
<tr>
<td>World Intellectual Property Organization</td>
<td>216 (1.37)</td>
</tr>
<tr>
<td>Deutsches Patent- und Markenamt</td>
<td>172 (1.09)</td>
</tr>
<tr>
<td>Canadian Intellectual Property Office</td>
<td>94 (0.6)</td>
</tr>
<tr>
<td>Spanish Patent and Trademark Office</td>
<td>46 (0.29)</td>
</tr>
</tbody>
</table>
A total of 99.73% (15,721/15,763) of the published patents originated from the top 10 IPOs, as detailed in Table 1. The China National Intellectual Property Administration (CNIPA) held the majority share, contributing 46% (n=15,763) published patents, surpassing other IPOs. The United States Patent and Trademark Office (USPTO) secured the second position with 25.71% (4052/15,763), followed by the Japanese patent office (JPO) at 13.79% (2174/15,763), the European Patent Office at 7.88% (1242/15,763), IP Australia at 1.55% (244/15,763), Inspirations Property Institution at 1.45% (228/15,763), and the World Intellectual Property Organization at 1.37% (216/15,763).

A total of 95.86% (15,110/15,763) of the published patents originated from the top 10 applicant countries, as illustrated in Figure 3. Among these, China emerged as the leading country, submitting the highest number of patents at 6540 (41.5%) of the total 15,763 publications. The United States secured the second position, contributing 23.69% (3734/15,763) of the publications, followed by Japan with 17.9% (2822/15,763).

The alignment of the top 3 IPOs with the leading patent applicant countries, namely, China, the United States, and Japan, underscores these regions as highly competitive in safeguarding technological innovations within the realm of digital health. This correlation suggests that institutions, enterprises, and individuals from these countries are at the forefront of technological innovation in the field of digital health. China, the United States, and Japan collectively stand out as leaders in driving advancements and securing intellectual property protection in the dynamic landscape of digital health technologies.

Multimedia Appendix 1 provides an overview of the nationality distribution of patent applicants within the top 3 IPOs. Specifically, 92.72% (2001/2158) of patent applicants to the JPO were from Japan, with 4.45% (96/2158) and 0.65% (14/2158) of applicants originating from the United States and South Korea, respectively. For the CNIPA, 87.08% (6316/7253) of patent applicants were from China, whereas 3.38% (245/7253) and 3.25% (236/7253) were from Japan and the United States, respectively. In the case of the USPTO, 69.32% (2953/4260) of patent applicants were from the United States, with 6.1% (269/4260) and 3.87% (165/4260) from Japan and China, respectively. Notably, the JPO had the highest proportion of domestic applicants, closely followed by the CNIPA. Japanese patentees, who also held the second-largest share of granted patents in China and the United States, actively expanded their digital health patents globally. Chinese patentees, despite ranking third based on the nationality of patent applicants in JPO and USPTO, exhibited a notable gap in percentage compared to their American and Japanese counterparts.

Institution Distribution
A total of 671 institutions or individuals were identified as publishers of patents related to digital health. As detailed in Table 2, Koninklijke Philips emerged as the leading enterprise with 329 published patents, followed by Siemens Healthcare GmbH (252 publications), Shanghai United Imaging Healthcare (151 publications), Samsung Electronics (121 publications), and Canon Medical Systems (118 publications). Notably, all these published patents were granted, with Koninklijke Philips having the highest number of granted patents. The top 10 institution types were exclusively enterprises, with Japanese corporations leading the count at 339 published patents, followed by American companies with 211 published patents, as indicated in Table 2.
**Table 2.** Top 10 institutions with the most published patents (N=15,763).

<table>
<thead>
<tr>
<th>Institution</th>
<th>Country of origin</th>
<th>Published patents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koninklijke Philips</td>
<td>The Netherlands</td>
<td>329 (2.09)</td>
</tr>
<tr>
<td>NV Siemens Healthcare GmbH</td>
<td>Germany</td>
<td>252 (1.6)</td>
</tr>
<tr>
<td>Shanghai United Imaging Healthcare</td>
<td>China</td>
<td>151 (0.96)</td>
</tr>
<tr>
<td>Samsung Electronics Co Ltd</td>
<td>South Korea</td>
<td>121 (0.77)</td>
</tr>
<tr>
<td>Canon Medical Systems Corp</td>
<td>Japan</td>
<td>118 (0.75)</td>
</tr>
<tr>
<td>Fujifilm Corp</td>
<td>Japan</td>
<td>115 (0.73)</td>
</tr>
<tr>
<td>Hitachi Ltd</td>
<td>Japan</td>
<td>106 (0.67)</td>
</tr>
<tr>
<td>International Business</td>
<td>United States</td>
<td>106 (0.67)</td>
</tr>
<tr>
<td>General Electric Company</td>
<td>United States</td>
<td>105 (0.67)</td>
</tr>
<tr>
<td>Biosense Webster (Israel) Ltd</td>
<td>Israel</td>
<td>97  (0.62)</td>
</tr>
</tbody>
</table>

**Table 3 highlights the top 10 highly cited institutions, all of which are enterprises.** Auris Health Inc and Masimo Corporation stand out with 849 and 983 citations globally, respectively, underscoring their substantial technological influence in the digital health field. It is noteworthy that Lepu (Beijing) Medical Equipment, founded in 1999, is the only Chinese company listed in **Table 3**.

**Table 3.** Top 10 highly cited institutes in the world.

<table>
<thead>
<tr>
<th>Institute</th>
<th>Cited frequency, n</th>
<th>Patents, n</th>
<th>Proportion (cited frequency/patents)</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auris Health Inc</td>
<td>849</td>
<td>19</td>
<td>44.68</td>
<td>United States</td>
</tr>
<tr>
<td>MASIMO corporation</td>
<td>983</td>
<td>28</td>
<td>35.11</td>
<td>United States</td>
</tr>
<tr>
<td>Ethicon Llc</td>
<td>700</td>
<td>24</td>
<td>29.17</td>
<td>United States</td>
</tr>
<tr>
<td>BERTEC Corporation</td>
<td>251</td>
<td>10</td>
<td>25.10</td>
<td>United States</td>
</tr>
<tr>
<td>VIGNET Incorporated</td>
<td>70</td>
<td>5</td>
<td>14.00</td>
<td>United States</td>
</tr>
<tr>
<td>Bardy Diagnostics Inc</td>
<td>60</td>
<td>5</td>
<td>12.00</td>
<td>United States</td>
</tr>
<tr>
<td>Lepu (Beijing)</td>
<td>149</td>
<td>13</td>
<td>11.46</td>
<td>China</td>
</tr>
<tr>
<td>Medical Hi LLC</td>
<td>69</td>
<td>7</td>
<td>9.86</td>
<td>United States</td>
</tr>
<tr>
<td>Amazon Technologies</td>
<td>74</td>
<td>8</td>
<td>9.25</td>
<td>United States</td>
</tr>
<tr>
<td>Heartflow Inc</td>
<td>43</td>
<td>6</td>
<td>7.17</td>
<td>United States</td>
</tr>
</tbody>
</table>

The analysis of coauthorship among institutions revealed noteworthy insights, as depicted in **Figure 4**. CiteSpace autonomously identified 215 institutions with 212 links, constructing a network illustrating institutional cooperation. The visual representation in **Figure 4** incorporates color-coded circles denoting the publication year, node size representing the quantity of published patents, annual-ring width indicating the number of published patents in each year, and link thickness reflecting cooperation strength. Notably, **Figure 4** highlights limited international cooperation in the field of digital health among institutions. However, the analysis identifies 2 notable enterprise alliances in Japan. The first alliance involves Fuji Film Corp, whereas the second features Panasonic Intellectual Property Management Corporation collaborating with both Fuji Film and Hitachi. Another noteworthy collaboration in Japan involves Canon, with Olympus maintaining a partnership, albeit with a lower link strength. These collaborative efforts signify strategic partnerships within the Japanese digital health landscape.
Conversely, there were relatively few local corporations collaborating with entities in the United States. General Electric Company exhibited robust research and development capabilities, evident by its limited linkages with other institutions, including universities.

In China, extensive collaboration was observed among hospitals, universities, and research institutions, providing substantial technological support to Shanghai United Imaging Healthcare. This collaborative landscape underscores the active engagement of various entities in China’s digital health sector, fostering advancements through cooperative efforts. Multimedia Appendix 2 illustrates the collaboration direction between Shanghai United Imaging Healthcare, the sole Chinese company listed in Table 3, and Zhongshan Hospital, its primary joint patent application agency, depicted through a word cloud. The primary focus of collaboration between these 2 entities centered around medical imaging, particularly in the fields of tumors and the spine. This collaboration highlights the synergy between Shanghai United Imaging Healthcare, and Zhongshan Hospital in advancing medical imaging technologies, with a specific emphasis on tumor and spine-related applications.

**Inventor Distribution**

Table 4 outlines the top 10 most active inventors based on patent filings, featuring 7 Americans and 3 Chinese innovators. This compilation was conducted using Microsoft Excel 2016. Notably, Assaf Govari secured the top position with 43 (0.3%) patents of the total 15,763. Frederick E published 31 (0.2%) patents of the total 15,763. Ammar Al Ali, Gary A, Jason L, Yong Wang, and Hairong Zheng collectively published 25 (0.2%) patents of the total 15,763, earning them the third spot in terms of publications. Among the top 10 inventors, 1 hailed from Zhejiang University, the remaining 9 were all affiliated with enterprises, highlighting the significant contributions of both academic and corporate entities in the realm of digital health innovation.

<table>
<thead>
<tr>
<th>The first inventor$^a$</th>
<th>Nationality</th>
<th>Institution</th>
<th>Published patents, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assaf</td>
<td>Israel</td>
<td>Biosense Webster (Israel) Ltd</td>
<td>43</td>
</tr>
<tr>
<td>Frederick E</td>
<td>United States</td>
<td>Ethicon Llc</td>
<td>31</td>
</tr>
<tr>
<td>Ammar Al</td>
<td>United States</td>
<td>Masimo Corporation</td>
<td>25</td>
</tr>
<tr>
<td>Gary A</td>
<td>United States</td>
<td>Zoll Medical Corporation</td>
<td>25</td>
</tr>
<tr>
<td>Jason L</td>
<td>United States</td>
<td>Ethicon Llc</td>
<td>25</td>
</tr>
<tr>
<td>Yong Wang</td>
<td>China</td>
<td>Chison Medical Technologies Co Ltd</td>
<td>25</td>
</tr>
<tr>
<td>Hairong Zheng</td>
<td>China</td>
<td>Shenzhen National Res Institute of High-Performance Medical Devices Co Ltd</td>
<td>25</td>
</tr>
<tr>
<td>Gust H</td>
<td>United States</td>
<td>Bardy Diagnostics Inc</td>
<td>24</td>
</tr>
<tr>
<td>Avi</td>
<td>United States</td>
<td>Align Technology Inc</td>
<td>23</td>
</tr>
<tr>
<td>Tao Liu</td>
<td>China</td>
<td>Zhejiang University</td>
<td>20</td>
</tr>
</tbody>
</table>

$^a$Inventors publishing the same number of patents were sorted alphabetically.
Assaf Govari, a prominent inventor, served as a fellow in research and development at Biosense Webster, a leading international company specializing in diagnosing and treating heart rhythm disorders [27]. A comprehensive search in the incoPat Patent Data System revealed 565 patent applications listing Assaf Govari as one of the inventors. Figure 5 depicts the analysis of technology themes using the patent 3D sandbox. This innovative visualization represents the competitive landscape of technologies in a 3D topographic map, where peaks signify technology-intensive areas and troughs represent technology gaps. Each dot represents a patent, and proximity indicates relevance. The most frequently mentioned technology theme was data point location tracking for analyzing magnetic resonance imaging data (176 patents). The second most frequent theme was electroporation technology used in ablation procedures during cardiac surgery (173 patents). The third most frequent theme involved an inflatable balloon (138 patents). These insights provide a glimpse into the diverse and impactful contributions of Assaf Govari in advancing technological frontiers in the field of digital health.

Figure 5. Top 10 technology topics of Assaf Govari.

Ammar Al Ali held a position as a fellow at Masimo Corporation, a global health care technology enterprise renowned for its research and development of cutting-edge noninvasive patient monitoring techniques. Masimo Corporation also secured the second position among the top 10 leading institutes globally, as indicated in Table 3. Frederick E and Jason L Harris were affiliated with Ethicon Llc, ranking third among the top 10 institutions. Ethicon Llc is widely recognized for its research and development in advanced tissue management. Gust H Bardy served as the founder and chief medical officer of Bardy Diagnostics, a company that held the sixth position among the top 10 institutes globally. Bardy Diagnostics specialized in the development of heart monitors and arrhythmia detection devices, emphasizing the delivery of diagnostically accurate and patient-friendly heart patches and other monitoring solutions. Gust H Bardy oversaw all clinical services related to ambulatory cardiac monitoring within the company.

Our analysis extended to coauthorship, using CiteSpace to scrutinize inventor networks. The software meticulously tallied all inventors in each patent, resulting in a coauthorship network of 168 inventors, 639 linkages, and a density of 0.05. Figure 6 visually represents this coauthorship network, with colors of circles denoting the publication year, node size indicating the quantity of published patents, annual-ring width reflecting the number of published patents in each year, and link thickness conveying cooperation strength. Figure 6 reveals that most inventors published patents in 2016 or 2017, and a decline is observed in the number of inventors contributing to technology innovation over subsequent years. This temporal trend offers
insights into the dynamics of inventor collaboration, emphasizing the early surge in collaborative efforts in the field of digital health innovation.

Figure 6. Inventor coauthorship network.

Centrality serves as an indicator of a node’s essentiality within a system. CiteSpace uses this metric to discern the relevance of authors or institutions, highlighting those with a centrality of at least 0.1 using a purple circle. In our analysis, Wang X emerged as a highlighted node, and we identified 142 published patents listing Wang X as the inventor.

Technological Topic and Patents With a High Citation Number

We conducted an investigation into the technological themes of digital health patents using the IPC Code and the DMC. Our searches and categorizations of patent materials were guided by hierarchical classification systems. The IPC, as the sole global classification system for patent documents, serves as a valuable tool for systematically organizing patents, establishing a foundation for the selective release of information, and providing a starting point for research into the state of the art in specific technological fields. The seventh edition of the IPC comprises 8 parts, divided into 120 classes, 628 subclasses, and approximately 69,000 groups [28]. Data for these research categories were extracted from search results in the Derwent Innovation Index database system and the incoPat system.

Table 5 delineates the foremost 10 technology domains within the realm of digital health, as categorized by DMC and IPC numbers. It is noteworthy that a single patent may fall under multiple IPC and DMC classifications, thereby impacting the overall count of patents. Of particular significance in the classification system of DMC were “diagnostic devices” (9081/15,763, 57.61%) and “claimed software products” (3781/15,763, 23.99%), which emerged as the 2 paramount technology subjects in the landscape of digital health patents. In addition, “medical equipment and information systems” (14,579/15,763, 92.49%) played a pivotal role in the overall categorization. Within the spectrum of coding, T01-J (data processing system), P31-A (diagnosis or surgery apparatus), and S05-D (electrical diagnosis) predominated as the primary codes for the sample patents. Notably, the special code listed was B11-C11, denoting “general computing methods and apparatus,” a prevalent technique within the realm of computer science. This underlines its ubiquity and relevance in the digital health patent landscape.
Table 5. Top 10 technology topics classified by Derwent Manual Code and International Patent Classification.

<table>
<thead>
<tr>
<th>Content</th>
<th>Published patents, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Derwent Manual Code</strong></td>
<td></td>
</tr>
<tr>
<td>T01-J06A</td>
<td>Medical equipment and information systems</td>
</tr>
<tr>
<td>P31-A05</td>
<td>Diagnostic devices</td>
</tr>
<tr>
<td>T01-S03</td>
<td>Claimed software products</td>
</tr>
<tr>
<td>T01-J10B2</td>
<td>Image analysis</td>
</tr>
<tr>
<td>S05-D</td>
<td>Electrical diagnosis</td>
</tr>
<tr>
<td>B11-C11</td>
<td>General computing methods and apparatus</td>
</tr>
<tr>
<td>S05-D01</td>
<td>Measuring and recording systems</td>
</tr>
<tr>
<td>T01-J10B1</td>
<td>Image enhancement</td>
</tr>
<tr>
<td>P34-A02</td>
<td>Syringes for removing and introducing fluids into the body</td>
</tr>
<tr>
<td>P31-A01</td>
<td>Surgical tools and instruments</td>
</tr>
<tr>
<td>S05-D07</td>
<td>Diagnostic displays and monitors</td>
</tr>
</tbody>
</table>

| **International Patent Classification number** | |
| A61B5 | Measurement for diagnostic purposes; human identification instruments used for radiation | 5962 |
| A61B6 | Diagnosis combined with radiotherapy equipment | 1769 |
| G06T7 | Image analysis | 1523 |
| G06K9 | Method or device for pattern recognition | 1214 |
| A61B34 | Manipulators or robots specially adapted to surgery; computer-assisted surgery | 957 |
| A61B8 | Diagnosis with an ultrasonic, acoustic, or infrasonic waves input device used to convert the data to be processed into a form that can be processed by the computer | 878 |
| G06F3 | Output device used to transfer data from the processor to the output device, for example, interface device equipment for testing eyes | 696 |
| A61B3 | Instrument for checking eyes electrotherapy | 690 |
| A61N1 | ICT* dedicated to arranging or managing health care resources or facilities | 659 |
| G16H40 | ICT dedicated to operating medical equipment or devices | 650 |

*aICT: information and communication technology.

From the perspective of the IPC number, pivotal metrics in the field of digital health were evident, with diagnostic and identification measurements (5962/15,763, 37.82%), instruments used for radiation diagnosis (1769/15,763, 11.22%), and image analysis (1532/15,763, 9.72%) emerging as the top 3 technology categories. Within the IPC classification, the sample patents displayed extensive distribution, notably within A61B (diagnosis, surgery, and identification) and G06 (computing and calculating or counting). A noteworthy inclusion in the coding spectrum was the special code G16H04, indicating that information and communication technology has been purposefully designed for the management or administration of health care facilities or resources or the operation of medical equipment and devices. This specialized code underscores the intentional convergence of technology and health care management within the digital health patent landscape.

CiteSpace served as the analytic tool to scrutinize category phrase bursts, facilitating a comprehensive exploration of temporal trends and hot spots within technological themes. Subject categories exhibiting robust burst strengths are indicative of heightened attention from the scientific community during specific periods. These subject categories, derived directly from the DMC of cited patents, encapsulate the primary areas of focus and exploration within a given field during a distinct timeframe. To map the landscape of subject categories, a cocitation network was constructed using the top 20 category themes identified each year. The extraction of the 20 most bursty category themes relied on CiteSpace’s burstiness findings, cataloged in Table 6. The “Year” column denotes the year when a code first appeared, whereas “Strength” quantifies the citation burst intensity for the listed codes. The sorting of these categories is based on the inception year, indicating when a category first exhibited bursts, and is detailed in the fourth column. The “End” column signifies the concluding year of a bursting category, providing a temporal context to the identified trends.
Table 6. Top 10 patents according to number of citations.

<table>
<thead>
<tr>
<th>Publication number</th>
<th>Institute</th>
<th>Year</th>
<th>Cited frequency, n</th>
<th>Client</th>
<th>Client-related product</th>
</tr>
</thead>
<tbody>
<tr>
<td>US10420551B2</td>
<td>Covidien Lp</td>
<td>2019</td>
<td>80</td>
<td>Physicians</td>
<td>Maryland elbow closed cutting surgical instrument</td>
</tr>
<tr>
<td>US10980535B2</td>
<td>Ethicon Llc</td>
<td>2021</td>
<td>77</td>
<td>Physicians</td>
<td>Ottava laparoscopic surgery robot</td>
</tr>
<tr>
<td>US10464209B2</td>
<td>Auris Health Inc</td>
<td>2019</td>
<td>63</td>
<td>Physicians</td>
<td>Auris robotic endoscopy system</td>
</tr>
<tr>
<td>US10539478B2</td>
<td>Auris Health Inc</td>
<td>2020</td>
<td>54</td>
<td>Physicians</td>
<td>Monarch robot endoscope platform</td>
</tr>
<tr>
<td>US10416264B2</td>
<td>Hyperfine Research Inc</td>
<td>2019</td>
<td>54</td>
<td>Physicians</td>
<td>Swoop portable MRI&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>US10426559B2</td>
<td>Auris Health Inc</td>
<td>2019</td>
<td>52</td>
<td>Physicians</td>
<td>Monarch robot endoscope platform</td>
</tr>
<tr>
<td>US10140544B1</td>
<td>12 Sigma Technologies</td>
<td>2018</td>
<td>50</td>
<td>Physicians</td>
<td>σ-Discover or Stroke CT&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>US10070799B2</td>
<td>Pison Technology Inc</td>
<td>2018</td>
<td>49</td>
<td>Patients</td>
<td>Pison AF neural insights</td>
</tr>
<tr>
<td>US10932729B2</td>
<td>Masimo Corporation</td>
<td>2021</td>
<td>47</td>
<td>Patients</td>
<td>Root platform</td>
</tr>
<tr>
<td>US10282914B1</td>
<td>Bao Tran; Ha Tran</td>
<td>2019</td>
<td>43</td>
<td>Physicians</td>
<td>VR&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>MRI: magnetic resonance imaging.
<sup>b</sup>CT: computed tomography.
<sup>c</sup>AI: artificial intelligence.
<sup>d</sup>VR: virtual reality.

As illustrated in Figure 7, the temporal evolution of digital health technology topics is graphically represented. The red line denotes the commencement year of each topic burst, whereas the subsequent timeline is delineated in green. Figure 7 showcases the top 20 DMC with the most robust burst strength in patents within the digital health domain, characterized by distinct thematic trajectories. The early phase of technological exploration predominantly encompassed topics such as “hospital equipment for patients’ medical records” and “general diagnostic image processing.” These themes garnered widespread citations from 2016 to 2017, exhibiting burst strengths of 33.171 and 24.483, respectively. This implies that patients’ medical records served as a primary data source, and general diagnostic image processing emerged as a principal application scenario for digital health technology during this period.
A subsequent thematic surge in digital health technology research occurred from 2017 to 2019, with a focus on establishing electronic image and data security. The second burst peak materialized in 2019 with the emergence of “medical IT system,” holding the strongest strength in the identified list, and its influence persisted until 2022. In addition, themes such as “knowledge processing by artificial intelligence with neural networks,” “medical equipment with the medical system,” and “medical information system for data processing” coincided with this burst period. Distinct trends unfolded from 2020 to 2022, with themes such as “sterilization and disinfection devices,” “electrical diagnosis for neurological currents and signals,” and “medical simulation systems” gaining prominence. The technology topics with the highest burst strength can be broadly categorized into 4 domains. First, the integration of medical equipment or apparatus with information technology systems, exemplified by themes such as “surgical tools,” “sterilization and disinfection devices,” “medical equipment with the medical system,” and “hospital equipment for patients’ medical records.” Second, themes related to medical information systems, such as “general diagnostic image processing,” “signal pattern recognition,” “forming electronic image,” and “medical simulation systems.” The third category encompasses methods for processing medical data, including themes such as “general diagnostic image processing,” “processing of recording image of radiation diagnosis using x-rays,” “electrical diagnosis for neurological currents and signals,” and “nonvehicle navigation.” Furthermore, key technology topics are intricately linked to data security and machine quality, as reflected in the themes of “program security management” and “computerized control of manufacturing of industrial machines and quality control.”

Table 6 outlines the top 10 most cited patents in the digital health field, categorized by the intended users—physicians and patients. Key features of these highly cited patents include a predominant focus on technologies benefiting physicians, with most patentees representing US-based corporations. The granted technologies are primarily method patents, featuring algorithms for tasks such as capturing real-world anatomy, detecting...
misalignment of robotic arms during surgery, and precise segmentation of medical digital images.

Among the top 10 patents, 4 (40%) prominently center around algorithmic technology. Notably, convolutional neural networks played a pivotal role in advancing the technological landscape. They significantly contributed to the development of image detection in the human brain through magnetic resonance [29] and played a key role in the accurate and efficient segmentation of medical images of human organs [30]. The patent US10426559B2, owned by Auris Health Inc, revealed a method for calibrating a medical instrument equipped with an articulable elongated shaft [31], showcasing advancements in instrument precision. In contrast, the patent titled “Systems and methods for computer-assisted operation” stood out as the sole patent applied by individuals. It unveiled a comprehensive technique for capturing real-world anatomy, implemented in 3D printing within the medical field [32].

Furthermore, 2 additional patents delve into the domain of robotic arms extensively used in surgical procedures. These patents are under the ownership of Auris Health Inc, a Subsidiary Corporation of Ethicon Llc. The patent US10464209B2 presents a system designed to control the position of manipulators both before and during medical procedures. It provides an illustrative example of this technology in action during a diagnostic “and” or “or” therapeutic bronchoscopy procedure [33]. Moreover, the patent addressing the “Detection of misalignment of robotic arms” unveils a system engineered for identifying undesirable forces on manipulators within a surgical machine system. This innovative technique is exemplified in the context of ureteroscopy and laparoscopic procedures [34].

In addition, 2 patents focus on technological advancements aimed at enhancing surgical instruments. The most widely cited patent describes a powered surgical instrument designed to enhance the reliability of communication between the disposable loading unit and the handle assembly. This patent, applied by Covidien Lp, a globally recognized enterprise in the medical field, was acquired by Medtronic in 2015 [35]. The patent with the identifier US10980535B2 introduces a surgical instrument tailored for use in endoscopic surgical cutting and fastening procedures [36]. Notably, this patent served to upgrade the Auris Robotic Endoscopy System, the first endoscopic robot developed by Auris Health Inc for the treatment of lung diseases. The system had successfully obtained approval from the US Food and Drug Administration in 2016. It is evident that these 2 highly cited techniques are product patents, conferring a broader scope of protected rights compared to method patents. This distinction underscores the significance of innovations in surgical instrument design and functionality within the realm of digital health.

Moreover, it is noteworthy that the technique titled “Detecting and using body tissue electrical signals,” owned by Pison Technology Inc, holds a prominent position in Table 6 and finds widespread implementation in wearable devices. Pison Technology Inc has articulated the versatile application of this technique across industrial, business, military, and medical domains. Functionally, it is designed for monitoring, control, feedback, actuation, communication, and comprehensive data accumulation and analysis. This innovative technique has practical applications such as monitoring heart rate and electrocardiograph from the wrist or foot. In addition, it is instrumental in measuring or analyzing fitness parameters, including muscular strength and stamina. In the medical field, it plays a crucial role in observing the state and progression of neurological disorders such as amyotrophic lateral sclerosis and other neurodegenerative conditions. The technique’s carrier is a wireless wrist-mounted user interface device, complemented by an app and a universal human-machine interface platform, as detailed in the patent [37]. This comprehensive approach reflects the broader impact and potential of digital health technologies, particularly in the realm of wearable devices and health monitoring.

The final patent to be highlighted is the one titled “Opioid overdose monitoring,” owned by Masimo Corporation. This innovative technique comprises an oximeter designed to be compatible with a handheld monitor for physiological parameters. The system is specifically engineered for monitoring indications of opioid overdose and facilitating the delivery of therapeutic drugs [38].

Discussion

Principal Findings

It is crucial to acknowledge that a patent represents an exclusive right granted by a nation to an institution or individual for an invention characterized by innovation, a technical solution, and industrial applicability. Patent information encapsulates a wealth of technological, commercial, and competitor knowledge, with a substantial portion often remaining unpublished. Although bibliometric analyses have been traditionally drawn from articles in digital health–related journals, this study marks the first systematic analysis of digital health technologies using patent documents. This approach serves as a pioneering effort in scientific knowledge dissemination, aiming to identify key nations, inventors, and technological focal points within the realm of digital health. The insights derived from this analysis are intended to provide a valuable reference for researchers and inventors navigating the dynamic landscape of digital health innovation.

In this study, our analysis of patent layouts across different countries revealed Japan as the most active nation in terms of overseas patent distribution, as per statistics from CNIPA, JPO, and USPTO. Notably, we identified distinct cooperation models prevalent in major countries. During the developmental stages of digital health, China, in contrast to the United States and Japan, appeared to be in a startup phase and exhibited a preference for collaborations involving universities, hospitals, and enterprises. In Japan, where digital health industrialization has reached a relatively mature stage, an intercompany cooperation model was found to be more suitable. In the United States, being the most advanced country in digital health industrialization, the need for cooperation seemed less essential, given the presence of formidable companies such as General
Electric Company, equipped with robust technological development capabilities.

Furthermore, our examination of patents allowed us to delineate the top 20 subject categories exhibiting the strongest citation bursts. This analysis serves to illuminate the technological development paths within the field of digital health, offering valuable insights into the trajectory of innovation in this domain.

The comparison with the latest bibliometric analysis on digital technologies in health care by Sikandar et al [39], which focused on published articles from 2017 to 2021, reveals interesting disparities in findings. Sikandar et al [39] identified the United Kingdom as the most active country in the research field of digital technology in health care. Noteworthy themes from the sampled articles included “Digital health literacy,” “Digital health for healthcare workers,” “Digital health and covid 19,” and “Applications of digital health” [39]. This comparison underscores a significant divergence in the hot spots and emphases between patents and articles within the digital health domain. While the United Kingdom emerged as the most active country in the article-centric analysis, our study, centered on patents, identified Japan as the most active nation. Furthermore, the themes that gained prominence in articles, such as “Digital health literacy,” were notably absent or had limited mention in the patent landscape. This discrepancy highlights the varied perspectives and priorities of authors and inventors, suggesting that the focus of researchers, as reflected in published articles, may not entirely align with the areas of innovation and patent activity in the digital health field.

The analysis of digital health patents spanning from 2017 to 2021 reveals a notable increase in granted patents in 2018, driven by several key factors. First, supportive policies played a pivotal role in promoting the implementation of digital technology in the medical field. Initiatives such as the “Framework for FDA’s Real World Evidence Program,” “Regulatory decision making for medical devices supported by real-world evidence,” and “Implementation of eHealth Records in Clinical Trial Guidance for Industry,” launched by the US Food and Drug Administration in 2017 and 2018, provided a regulatory framework and impetus for digital health innovation. Furthermore, the Chinese government’s strategic initiatives, as outlined in the “Guidance on Promoting and Regulating the Medical Data Applications” and “Regulations on promoting the development of online healthcare,” released by the General Office of the Chinese State Council in 2016 and 2018, respectively, significantly contributed to the growth of digital health in China. The introduction of the General Data Protection Regulation by the European Union on May 25, 2018, marked a milestone in the formation of industry standards for digital health. This regulation, applicable to any entity handling personal data related to European Union member states, had a global impact on data protection practices. Moreover, the partial coverage of digital health costs by medical insurance in some countries, exemplified by the development of a digital formulary in the United States in 2017, enhanced accessibility and financial viability. This, in turn, stimulated innovation and contributed to the surge in granted patents. In summary, a convergence of regulatory support, strategic government initiatives, the establishment of industry standards, and advancements in insurance coverage collectively fueled the remarkable increase in digital health patents in 2018.

There is no patent layout for digital health in Africa, and possible reasons are as follows. Policy guidance, technological innovation, and financial support are the foundation on which digital health is born and grows. Africa lacks the corresponding economic basis and talent reserve for the development of the digital health. Meanwhile, few corporation giants focusing on digital health are registered in this area because of the turbulent politics situation, poor economic foundation, and low education level, and this area also has no more mature commercial application for the digital health technology.

This study reveals the top countries, key inventors, and hot technologies in the patent landscape for digital therapies, which provides guidance for implementers who are applying for international patents on digital health. Innovation policies for digital health should be formulated and implemented in countries where patents are predominantly applied. The key inventors and hot technologies proposed by the study can be used as both a reference and as an inspiration for inventors to find cooperation partners and pioneer new research and development areas in the near future.

There is no established patent layout for digital health in Africa, and several factors contribute to this absence. The growth of digital health relies on a foundation of policy guidance, technological innovation, and financial support. Unfortunately, Africa lacks the corresponding economic basis and talent reserve necessary for the development of digital health. In addition, the turbulent political situation, poor economic foundation, and low education levels in the region deter major corporations from focusing on digital health, and there is a lack of mature commercial applications for digital health technology. In addition, it is possible that our search methods may impose limitations on the retrieval of patents from the African region.

This study not only underscores the challenges but also reveals key insights into the top countries, key inventors, and emerging technologies in the global patent landscape for digital therapies. These findings offer valuable guidance for international patent applicants in the field of digital health. Countries where patents are predominantly applied should consider formulating and implementing innovation policies for digital health. Moreover, the key inventors and emerging technologies identified in this study can serve as both a reference and an inspiration for inventors. They provide a roadmap for finding potential cooperation partners and exploring new research and development areas in the near future.

Limitations

Nevertheless, our study has certain limitations. First, we retrieved only a partial data set from the Derwent Innovations Index database, and some databases were not included. This omission may have resulted in overlooking certain frontier technologies. However, it is worth noting that the Derwent Innovations Index database, from which we collected most published patents, encompasses patents from >50 patent issuing authorities worldwide. Second, our analysis focused solely on granted and valid patents, emphasizing right stability. This

https://www.i-jmr.org/2024/1/e48259
(page number not for citation purposes)
approach may overlook some significant patents in the initial application stage. Third, our study was dedicated to analyzing the technical construction of digital health patents through bibliometric analysis. We did not provide a detailed discussion on the content of patents. For a more comprehensive understanding, a systematic patent review in the future is essential. For future work, expanding the scope of patent searches by including searches on Google Scholar could enrich and fill gaps in patents that were not investigated in this study.

Conclusions

We conducted an analysis of the digital health technology trend from 2017 to 2021 using a data set of 15,763 published patents extracted from the Derwent Innovations Index database. The study not only identified the most productive countries, institutions, and inventors but also delineated the various stages of technology development, key technology categories, and the most cited patents within the digital health domain.

In terms of patent participation rates, China emerged as a frontrunner among the top 10 countries or regions, leading in both the number of patents and technological advancements. This dominance is attributed to China’s concerted efforts in driving digital transformation across various sectors, particularly in medicine. In addition, the substantial patient populations in China contribute significantly to the wealth of data resources available for advancing digital health technologies.

Among the top 10 institutions with the highest number of published patents and the top 10 highly cited institutes globally, all were enterprises. This underscores the pivotal role that corporations play in shaping the patent landscape in the digital health sector. The active engagement of these enterprises is anticipated to yield considerable economic returns in the near future.

Examining the top 10 inventors featured in each patent list, 6 hailed from the United States. Furthermore, the patentees of the leading 10 patents, based on the number of citations, were all Americans. Overall, 90% (9/10) of the highly cited institutes globally also originated from the United States. Undoubtedly, the United States has been a frontrunner in the digital health technology domain. Illustrating a noteworthy collaboration model is Shanghai United Imaging Healthcare, a leader in advanced health care imaging techniques. The company closely collaborated with Zhongshan Hospital, a renowned health institution in the cardiovascular field in China, in the development of digital health solutions.

In our analysis of highly cited patents, we observed that health interventions and improvements in surgical devices were the primary application scenarios for digital health technology. Core to these advancements were algorithms, the focal point of patent-protected technologies. In addition, technologies related to data transfer, storage, and exchange, particularly in the context of telehealth, are anticipated to be hot spots in basic research in the near future.

Acknowledgments

This work received support from the Intellectual Property Operation Center of the Shanghai Healthcare System (JL), the National Nature Science Foundation of China (grant 71704136 to WG), Shanghai Science and Technology Innovation Action Plan of Building a Regional Innovation Community in the Yangtze River Delta (grant 20642430100 to JL), and the Shanghai Engineering Technology Research Center and Professional Service Platform Construction Project (grant 20DZ2293800 to JL). The funders were not involved in study design, data collection and analysis, decision to publish, or manuscript preparation.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

WG contributed to formal analysis and original draft writing. JW contributed to formal analysis. SL and YZ handled image analysis. SL, ZA, and JL played a key role in conceptualization, funding acquisition, and both original draft writing, review, and editing. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1


[ PNG File, 238 KB - ijmr_v13i1e48259_app1.png ]

Multimedia Appendix 2

Words cloud of the collaboration between Shanghai United Imaging Healthcare, and Zhongshan Hospital.

[ PNG File, 133 KB - ijmr_v13i1e48259_app2.png ]
References


2. Digital therapeutics alliance homepage. Digital Therapeutics Alliance. URL: https://dtxalliance.org/ [accessed 2021-08-09]


10. Therapeutic areas for which digital therapeutic products are available or currently being developed. Digital Therapeutics Alliance. 2020. URL: https://dtxalliance.org/wp-content/uploads/2020/03/DTx-Disease-State-Targets_03.20.pdf [accessed 2021-09-10]


Abbreviations

- CNIPA: China National Intellectual Property Administration
- DMC: Derwent Manual Code
- DTA: Digital Therapeutics Alliance
- IPC: International Patent Classification
- IFO: intellectual property office
- JPO: Japanese patent office
- USPTO: United States Patent and Trademark Office

©Wenjun Gu, Jinhua Wang, Yunqi Zhang, Shaolin Liang, Zisheng Ai, Jiyu Li. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 11.07.2024. 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.
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.
Behavioral Insights from Vaccine Adoption in Nigeria: Cross-Sectional Survey Findings

Sohail Agha1,2, BA, PhD; Ifeanyi Nsorfor3, MSc, MD; Drew Bernard2, BA; Sarah Francis2, BA; Nandan Rao4, MSc

1Behavior Design Lab, Stanford University, Stanford, CA, United States
2Behavioral Insights Lab, Seattle, WA, United States
3Centre for Family Health Initiative, Abuja, Nigeria
4Virtual Labs, Corvallis, OR, United States

Corresponding Author:
Sohail Agha, BA, PhD
Behavior Design Lab
Stanford University
581 Capistrano Way
Stanford, CA, 94305
United States
Phone: 1 1 206 351 9346
Email: sohailagha@gmail.com

Abstract

Background: To generate behavioral insights for the development of effective vaccination interventions, we need approaches that combine rapid and inexpensive survey data collection with instruments based on easy-to-use behavior models. This study demonstrates how an inexpensive digital survey helped identify the drivers of COVID-19 vaccination in Nigeria.

Objective: This study aims to illustrate how behavioral insights can be generated through inexpensive digital surveys.

Methods: We designed and conducted a cross-sectional survey with multistage sampling. Data were collected from Nigerians (aged ≥ 18 years) from 120 strata based on age, sex, state, and urban or rural location. Respondents were recruited via advertisements on Meta platforms (Facebook and Instagram) using the Virtual Lab open-source tool. We used a Meta Messenger chatbot for data collection; participants were compensated with 400 naira (US $0.87 cents). Data collection took 2 weeks. In total, 957 respondents completed the survey, at an advertising cost of US $1.55 per respondent. An 18-item instrument measuring core motivators, ability barriers, sociodemographic characteristics, and respondents’ vaccination status was pretested before data collection. We ran separate logistic regression models to examine the relationships between vaccine uptake and core motivators, ability barriers, and sociodemographic variables. A final model that predicted vaccine uptake included all 3 sets of variables.

Results: About 56% (n=540) of respondents reported that they had received at least 1 COVID-19 vaccination. Three core motivators were positively associated with vaccine uptake: the belief that the COVID-19 vaccine promised a better life (adjusted odds ratio [aOR] 3.51, 95% CI 2.23-5.52), the belief that the vaccine would allow respondents to do more things they enjoyed (aOR 1.97, 95% CI 1.33-2.93), and respondents’ perception that their friends and family members accepted their decision to get vaccinated (aOR 1.62, 95% CI 1.06-2.48). Two ability barriers were negatively associated with vaccine uptake: cost- or income-related concerns lowered the odds of being vaccinated (aOR 0.35, 95% CI 0.24-0.50) and the lack of availability of vaccines at places respondents routinely visited also lowered their odds of being vaccinated (aOR 0.29, 95% CI 0.21-0.40). After adjusting for other variables, the perceived fear of getting COVID-19 and the hardship associated with the disease were no longer associated with vaccine uptake.

Conclusions: These findings suggest that hope is more important for Nigerians than fear when it comes to vaccine adoption, enjoying life is more important than worrying about getting the disease, and approval from friends and family is more powerful than their disapproval. These findings suggest that emphasizing the benefits of leading a fuller life after being vaccinated is more likely to succeed than increasing Nigerians’ fear of COVID-19. This study identifies a very different set of factors associated with COVID-19 vaccine adoption than previous Nigerian studies.

Introduction

Background

Globally, the COVID-19 pandemic has highlighted the importance of behavioral insights for increasing the use of preventive behaviors such as wearing a mask, social distancing, and getting vaccinated. However, recent studies on COVID-19 vaccine acceptance emerging from Nigeria and other low- and middle-income countries (LMICs) have primarily focused on identifying gaps in knowledge, attitudes, and beliefs associated with vaccine hesitancy. A number of these studies recommend educating health care workers (HCWs) and members of the general population on vaccine safety and efficacy and assume that equipping people with factually correct information will allay their concerns, increase their perceived risk of acquiring COVID-19, and lead to higher rates of vaccine adoption [1,2].

Data from other Nigerian studies, however, raise questions regarding the strength of the relationship between risk perception and vaccine acceptance. A recent hospital-based study in southern Nigeria found that, while most HCWs perceived themselves at risk of COVID-19, only about half were willing to receive the COVID-19 vaccine [3]. Substantial gaps between risk perception and willingness to accept a COVID-19 vaccine have been observed in the general population in northern Nigeria as well [4]. A study that interviewed over 5000 respondents across all states in Nigeria found that COVID-19 was not perceived as a threat by most respondents [5].

What seems to be an important determinant of vaccine acceptability in Nigeria is trust in the vaccine manufacturing process, health system, government [6], and institutions involved in risk communication on behalf of the government [7]. Yet, despite multiple studies showing a weak relationship between risk perception and willingness to adopt a COVID-19 vaccine in Nigeria, researchers continue to recommend the provision of factually correct information to fill “information gaps” [4].

A recent systematic review of the COVID-19 literature in Nigeria shows that vaccination rates among those at high risk of COVID-19, such as HCWs, were lower than among those at low risk of COVID-19 [8]. Given the weak relationship between risk perception and vaccine uptake in Nigeria, it is not surprising that a recent evaluation found that risk communication efforts in Nigeria were inadequate in sustaining changes in behavior observed at the beginning of the pandemic [9]. Lawal [9] showed that during the first 30 days of the discovery of COVID-19 in Nigeria, and until the national lockdown, public interest in learning about the disease surged. Visits to public places such as grocery stores declined during this period as stringent government policies resulted in reduced mobility of the population. The study by Lawal [9] found that, as the Nigerian population started becoming aware of the disease, there was a slight decline in the number of COVID-19 cases. However, this decline occurred for a relatively short period of time. The number of new cases started increasing again as the initial effects of risk communication interventions dissipated. Lawal [9] concluded that Nigerians listened to messages telling them to take preventive measures such as social distancing or masking for some time but eventually got tired of the messages and stopped responding to them. In part, this was because the recommended public health precautions did not fit well in the context in which they lived their lives [9].

A clear picture of the drivers of COVID-19 vaccine hesitancy does not emerge from the recent public health literature on Nigeria, in part because much of this work is not based on a behavioral framework. The importance of theory-based work to understand the drivers of vaccine acceptance and design appropriate interventions has been emphasized [10]. In the absence of a clearly articulated framework for understanding vaccine-related behavior, it is difficult to interpret the findings of individual studies and arrive at a clear picture of what drives vaccine uptake in Nigeria.

As a result, there is very limited guidance available to support Nigerian practitioners in designing interventions that might accelerate COVID-19 vaccine uptake. For example, although many recent studies emphasize the importance of implementing health promotion interventions or increasing HCWs’ ability to communicate more effectively with members of the general public, most of these studies do not provide any guidance on what the content of this communication should be or what strategy should be used to persuade adults to get vaccinated. Thus, the available research is at a standstill in terms of providing insights that would help in designing more effective behavioral interventions to accelerate vaccine uptake in Nigeria.

An important reason that the literature does not provide a clear direction for the design of behavioral interventions is the lack of use of behavioral frameworks in explaining vaccine acceptance and uptake. Of the more than 20 peer-reviewed publications reviewed for this paper, we found only 1 that used a behavior model to interpret its findings [5]. This is not surprising as researchers have noted the limited use of behavioral frameworks in public health research and practice for over a decade [11-14].

The need for a behavior model that can be used to explain the vaccine adoption process in simple terms that resonate with practitioners is urgent. While a broader discussion of what a practitioner-friendly behavior model should comprise of is merited, the characteristics of such a model have been proposed [15]. A minimum criterion should be that the use of the model leads to deliberate programmatic decisions, a greater emphasis on strengthening activities supported by behavioral research findings, and the elimination of activities that are not evidence based.

The practitioner-friendly model used in this study, the Fogg Behavior Model (FBM), was introduced in the public health literature in 2019 to explain the effects of a social marketing behavior change campaign on the adoption of condoms by married men in Pakistan [13]. More recently, it has been used to identify behavioral drivers associated with the (1) adoption...
of COVID-19 vaccination by a low-income population in Cote d’Ivoire [16], (2) adoption of iron folate by pregnant women in India [15], (3) uptake of COVID-19 vaccination by HCWs in Nigeria [17,18], and (4) use of contraception by adolescent girls and young women in Nigeria [19]. A recent study also demonstrates the use of the FBM in making timely programmatic adjustments to a contraceptive social marketing intervention implemented in Nigeria [20]. To the best of our knowledge, this is the first time that the FBM is being applied to understand the dynamics of COVID-19 vaccine adoption in the general population of Nigeria.

FBM: Motivation and Ability as the Drivers of Behavior

The FBM is a model developed for use by practitioners to understand the drivers of human behavior and assist them in the design of behavior change interventions. Fogg states that behavior happens when motivation, ability, and a prompt happen at the same moment. As shown in Figure 1, the model can be visualized in 2 dimensions. Figure 1 shows motivation along the y-axis and ability along the x-axis. Motivation ranges from high to low for any behavior. Ability also ranges from high to low for any behavior. For simplicity, we describe a behavior as being easy to do or hard to do. For a prompt to work, a person needs to have sufficient motivation and ability. The motivation-ability threshold is reflected by an action line in the FBM. Behavior occurs when a person whose motivation and ability are above the action line is prompted. The prompt does not work if the person does not have sufficient motivation to undertake the behavior and finds the behavior hard to do, that is, they are below the action line [21].

Methods

Questionnaire Design: Core Motivators

Fogg defines motivation as having 3 components: anticipation, sensation, and belonging. Anticipation reflects the hopes and fears a person associates with a behavior. Sensation reflects the pleasure or pain a person associates with a behavior. Belonging is reflected by the acceptance or rejection of the behavior by people whose opinions a person considers important.

We conducted a review of the literature to identify relevant constructs and appropriate measures of motivation and ability. The survey questionnaire (Multimedia Appendix 1) was designed using instruments developed to test motivation and ability constructs in the FBM. These instruments had been previously tested in the general population in Cote d’Ivoire and with HCWs in Nigeria and had shown acceptable levels of reliability [16,18]. Respondents were allowed to answer how strongly they agreed or disagreed with 6-point Likert scale items measuring core motivators, such as hope, fear, pleasure, pain, acceptance, and rejection, associated with the adoption of the COVID-19 vaccine.

Ability Barriers

Fogg defines ability in terms of 5 barriers: time, money, the physical effort required to adopt a behavior, the mental effort required to adopt a behavior, and whether the behavior fits into the person’s routine [21]. Fogg initially considered social
deviance (or social norms) as part of ability but did not include social norms as an ability barrier in later iterations of the model [22], possibly because of the complexity of the relationship between norms and behavior. To capture ability barriers, respondents could answer how strongly they agreed with items on a 5-point Likert scale. These items captured the 5 ability factors related to COVID-19 vaccine uptake: time, money, mental effort, physical effort, and routine. The instrument also contained questions on sociodemographic variables. In total, the instrument comprised 18 questionnaire items.

The instrument was pretested twice in Nigeria with samples of approximately 100 respondents. The first pretest showed that the relationship among variables measuring motivation, ability, and COVID-19 vaccine uptake was in the expected direction, with 1 exception. The variable measuring agreement or disagreement with the question “Many of my friends and family would think poorly of me if they knew I had taken the COVID-19 vaccine” did not demonstrate the expected relationship with vaccine uptake. This was replaced with the statement “Most people I know have obtained the COVID-19 vaccination.” The second pretest showed that the latter response was associated with the outcome in the expected direction. All other items included in the instrument demonstrated the expected relationships with vaccine uptake in both pretests.

Survey Design and Sampling

We implemented a cross-sectional survey with multistage sampling. Nigeria is a large and diverse country with 36 states and the Federal Capital Territory of Abuja. These states and the Federal Capital Territory are grouped into 6 geopolitical zones: northwest, northeast, north central, southwest, southeast, and south-south. Although this survey was not designed to be a representative survey of Nigeria, we aimed to capture the diversity of the country’s population by ensuring that respondents from 1 state within each geopolitical zone were sampled. Thus, the states were selected at the first stage. Sokoto state was selected from the northwest, Bauchi state from the northeast, Niger state from north central, Lagos state from the southwest, Anambra state from the southeast, and Rivers state from the south-south. The state selected within each zone reflects its socioeconomic, religious, and ethnic diversity. In addition, these 6 states had ongoing COVID-19 vaccination campaigns to help ensure that survey findings would be useful for COVID-19 program managers in Nigeria.

The sample was stratified by 6 states, 5 age groups, male and female sex, and urban or rural location. This resulted in 120 strata from which respondents were sampled. We ran a total of 120 different ad sets targeting respondents based on the characteristics mentioned above.

Respondents were recruited via advertisements on the Meta digital ad platforms (Facebook and Instagram) using the Virtual Lab open-source tool [23]. The Virtual Lab tool ran ads targeting respondents in all 120 strata. We used a Meta Messenger chatbot for the survey data collection, compensating respondents who completed the survey with 400 naira (US $0.87 cents) in mobile phone credit. Respondents who clicked on the ads were directed to the messenger bot. Of the 214,335 male and female respondents who reached through the ads, 3660 clicked on the link, 1367 started the survey, and 1011 answered most survey questions.

Respondents could complete the survey in one go or start the survey, stop, and return to complete it later. In total, 957 respondents answered all questions in the survey at an advertising cost of US $1.55 per person. Data from these 957 respondents are used for the analysis.

The outcome of interest for this study was having received at least 1 COVID-19 vaccination. The Government of Nigeria’s data on the number of vaccinations provided in 2022 suggested that 50% of Nigerians had received at least 1 COVID-19 vaccination. A sample size calculation was made using an outcome value of 50% and a design effect of 1.5. We estimated that a sample size of 900 would provide a margin of error of 4 percentage points on the outcome of interest.

Statistical Analysis: Relationships Among Core Motivators, Ability Barriers, and Vaccine Uptake

Univariate analysis was conducted to provide the frequency distributions of core motivators, ability barriers, and sociodemographic characteristics of the sample. Bivariate analysis was conducted to explore the relationships between core motivators, ability factors, sociodemographic characteristics, and having received at least 1 COVID-19 vaccination. Multivariate logistic regression analysis was conducted to identify which core motivators, ability factors, and sample characteristics had a significant relationship with vaccine uptake [24].

We also ran a final multivariate model to determine whether there was any change in the relationship between individual variables and vaccine uptake after taking the 3 sets of variables (core motivators, ability factors, and sociodemographic characteristics) into account. Adjusted odds ratios (aORs) from the analyses are shown in the tables. P values were considered statistically significant at P<.05.

Ethical Considerations

The ethical approval for the study was obtained from Nigeria’s National Health Research Ethics Committee (NHREC/01/01/2007). Informed consent was obtained from all respondents to the quantitative survey. Respondents were assured that all written and recorded data would be kept confidential by using codes to identify participants instead of names or any other personal identifiers. Respondents were informed about their right to refuse to participate in the study or withdraw at any time during the interview.

Results

Core Motivators and COVID-19 Vaccine Uptake

Table 1 shows the frequency distributions of core elements of motivation identified by the FBM, cross-tabulations between core motivators and COVID-19 vaccine uptake and the aOR of COVID-19 vaccination. The first column of Table 1 shows that about three-fourths (n=726, 75.9%) of survey respondents agreed or strongly agreed with the statement that getting vaccinated against COVID-19 allows a person to live a better life, a measure of hope. Nearly half (n=444, 46.4%) of respondents with
reported that getting vaccinated protected them and their families from hardship, a measure of fear. Nearly two-thirds (n=610, 63.7%) of respondents agreed or strongly agreed with the statement that getting vaccinated allows a person to do the things they enjoy, a measure of pleasure. About 81.5% (n=780) reported that vaccination reduces the likelihood of getting or spreading COVID-19, a measure of pain. About 79.1% (n=757) of respondents reported that many of their family and friends approve of the COVID-19 vaccination, a variable measuring social acceptance. Consistent with the high social acceptance of the COVID-19 vaccination, rejection of the vaccine was much lower, that is, 37.1% (n=355) of respondents reported that most of their family and friends did not approve of the COVID-19 vaccination.

Table 1. Frequency distributions of core motivators, cross-tabulations, and the adjusted odds of COVID-19 vaccination in Nigeria.

<table>
<thead>
<tr>
<th>Core motivators</th>
<th>Frequency distributions of core motivators (N=957, 100%), n (%)</th>
<th>Nigerian adults who obtained at least 1 vaccination (n=540, 56.4%), n (%)</th>
<th>P value</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hope: Vaccination allows a person to live a better life</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>726 (75.9)</td>
<td>483 (66.5)</td>
<td>3.40 (2.27-5.09)</td>
<td></td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>231 (24.1)</td>
<td>57 (24.7)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td><strong>Fear: Getting vaccinated protects people from hardship</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>444 (46.4)</td>
<td>301 (67.8)</td>
<td>1.09 (0.78-1.51)</td>
<td></td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>513 (53.6)</td>
<td>239 (46.6)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td><strong>Pleasure: Getting vaccinated allows people to do more things they enjoy</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>610 (63.7)</td>
<td>411 (67.4)</td>
<td>1.84 (1.30-2.62)</td>
<td></td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>347 (36.3)</td>
<td>129 (37.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pain: Getting vaccinated reduces the likelihood of getting or spreading COVID-19</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>780 (81.5)</td>
<td>477 (61.2)</td>
<td>1.15 (0.76-1.73)</td>
<td></td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>177 (18.5)</td>
<td>63 (35.6)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td><strong>Acceptance: Many friends and family approve vaccination</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>757 (79.1)</td>
<td>472 (62.4)</td>
<td>1.76 (1.21-2.58)</td>
<td></td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>200 (20.9)</td>
<td>68 (34)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td><strong>Rejection: Most family and friends do not approve of vaccination</strong></td>
<td></td>
<td></td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>355 (37.1)</td>
<td>344 (57.1)</td>
<td>1.06 (0.79-1.43)</td>
<td></td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>602 (62.9)</td>
<td>196 (35.2)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
</tbody>
</table>

²Pseudo R²=11.85%

The second column of Table 1 shows cross-tabulations between the core motivators of vaccination and vaccine uptake. There were large differences in vaccine uptake by core motivators at the bivariate level. Respondents who associated a COVID-19 vaccination with the hope of a better life had a 42-percentage point higher rate of vaccination (n=483, 66.5% vs n=57, 24.7%; P<.001). Those who feared the hardship that COVID-19 infection would bring had a 21-percentage point higher vaccination rate than other respondents (n=301, 67.8% vs n=239, 46.6%; P<.001). The pleasure that respondents associated with being able to do what they enjoyed doing because of being vaccinated was reflected by a 30-percentage point higher rate of vaccination (n=411, 67.4% vs n=129, 37.2%; P<.001). Those who believed that being vaccinated would result in getting or spreading COVID-19 was associated with a 25-percentage point higher rate of vaccination (n=477, 61.2% vs n=63, 35.6%; P<.001). Acceptance of the vaccine by friends and family was associated with a 28-percentage point higher vaccination rate (n=472, 62.4% vs n=68, 34%; P<.001). It is interesting that social rejection, or the lack of approval of the vaccination by family members, was not associated with vaccine uptake.

The third column of Table 1 shows the adjusted odds of COVID-19 vaccination. With all core motivators in the model, Nigerians who believed that COVID-19 vaccination was associated with a better life, had 3 times higher odds of getting vaccinated (aOR 3.40, 95% CI 2.27-5.09). Those who believed that getting vaccinated would allow them to do more things that they enjoyed were more likely to be vaccinated (aOR 1.84, 95% CI 1.30-2.62). Friends and family members’ acceptance of their vaccination was associated with a 1.76 times higher odds of getting vaccinated (aOR 1.76, 95% CI 1.21-2.58). It is interesting that social rejection, or the lack of approval of the vaccination by family members, was not associated with vaccine uptake.
decision to get vaccinated was associated with a higher vaccination rate (aOR 1.76, 95% CI 1.21-2.58).

### Ability Barriers and COVID-19 Vaccine Uptake

Table 2 shows the frequency distributions of ability factors identified by the FBM, cross-tabulations between ability factors and COVID-19 vaccine uptake, and the adjusted odds of COVID-19 vaccination. The first column of Table 2 shows that about 54% (n=516) of respondents felt that their family or work responsibilities made it difficult for them to get vaccinated. This variable measures the constraint of time. Over a third (n=347, 36.3%) of respondents felt that the cost or loss of income associated with getting vaccinated was a barrier. Nearly 40% (n=437, 46%) of the respondents felt that the decision to get vaccinated was difficult. This variable measures the mental effort required to get vaccinated. About 42% (n=401) of respondents reported that not having the vaccine available in places they routinely visited was a barrier to getting vaccinated. The latter measures the routine associated with adopting a behavior.

Table 2. Frequency distributions of ability factors, cross-tabulations, and the adjusted odds of COVID-19 vaccination in Nigeria.

<table>
<thead>
<tr>
<th>Ability Factor</th>
<th>Frequency Distributions of Ability Factors (N=957, 100%), n (%)</th>
<th>Nigerian adults who obtained at least one COVID-19 vaccination (n=540, 56.4%), n (%)</th>
<th>P value</th>
<th>Adjusted Odds of Obtaining at Least One COVID-19 Vaccination</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time: Family or work responsibilities make it difficult to find time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>516 (53.9)</td>
<td>277 (53.7)</td>
<td>.06</td>
<td>0.94 (0.69-1.27)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>441 (46.1)</td>
<td>263 (59.6)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td><strong>Money: Costs or loss of income make it difficult</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>347 (36.3)</td>
<td>158 (45.5)</td>
<td>&lt;.001</td>
<td>0.45 (0.33-0.62)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>610 (63.7)</td>
<td>382 (62.6)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td><strong>Physical effort: Physical effort makes it difficult</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>382 (39.9)</td>
<td>218 (57.1)</td>
<td>.74</td>
<td>1.46 (1.06-2.01)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>575 (60.1)</td>
<td>322 (56.0)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td><strong>Mental effort: Decision to get vaccine is difficult</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>437 (45.7)</td>
<td>237 (54.2)</td>
<td>.21</td>
<td>0.94 (0.70-1.27)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>520 (54.3)</td>
<td>303 (58.3)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td><strong>Routine: Vaccine not available where I routinely visit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree or strongly agree</td>
<td>401 (41.9)</td>
<td>141 (35.2)</td>
<td>&lt;.001</td>
<td>0.21 (0.16-0.28)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know</td>
<td>556 (58.1)</td>
<td>399 (71.8)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
</tbody>
</table>

The second column of Table 2 shows cross-tabulations between ability factors and vaccine uptake. Respondents who agreed or strongly agreed with the statement that costs or loss of income were a barrier reported a 17-percentage point lower vaccination rate compared to others (n=158, 45.5% vs n=382, 62.6%; P<.001). The lack of availability of the COVID-19 vaccine in places that they routinely visited was associated with a 37-percentage point lower rate of vaccination (n=141, 35.2% vs n=399, 71.8%; P<.001).

The third column of Table 2 shows the adjusted odds of a COVID-19 vaccination. With all ability factors in the model, Nigerians who believed that the cost or the loss of income made it difficult to obtain a COVID-19 vaccination were less likely to get vaccinated (aOR 0.45, 95% CI 0.33-0.62). The lack of availability of vaccines at places respondents routinely visited was associated with a lower likelihood of vaccination (aOR 0.21, 95% CI 0.16-0.28). Contrary to our expectations, respondents who felt that physical effort makes it difficult to get vaccinated were more likely to be vaccinated (aOR 1.46, 95% CI 1.06-2.01).

### Sociodemographic Factors and COVID-19 Vaccine Uptake

Table 3 shows the frequency distributions of sociodemographic characteristics of respondents in the sample, cross-tabulations between these characteristics and vaccine uptake, and the adjusted odds of COVID-19 vaccination. The first column of Table 3 shows that, as expected from a digital survey, the sample had relatively young participants: 56% (n=531) of respondents were aged 18-29 years and 16% (n=151) were aged 40 years and older. Male participants represented a higher proportion of the sample (n=592, 61.9%). About 39% (n=371) of respondents had a primary or secondary school certificate, one quarter
(n=231, 24.1%) had an ordinary national diploma (OND) or a higher national diploma (HND), and one-third of respondents (n=317, 33.4%) had a bachelor’s or higher degree. A majority of respondents were from urban areas: 57% (n=544) were from cities, 34% (n=323) from towns, and 9% (n=90) from rural areas.

Table 3. Frequency distributions of sociodemographic variables, cross-tabulations, and the adjusted odds of COVID-19 vaccination in Nigeria.

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Sample characteristics (N=957, 100%), n (%)</th>
<th>Nigerian adults who obtained at least 1 vaccination (n=540, 56.4%),</th>
<th>P value</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>531 (55.5)</td>
<td>300 (56.5)</td>
<td>.71</td>
<td>1.31 (0.89-1.93)</td>
</tr>
<tr>
<td>30-39</td>
<td>275 (28.7)</td>
<td>159 (57.8)</td>
<td></td>
<td>1.28 (0.85-1.93)</td>
</tr>
<tr>
<td>≥40</td>
<td>151 (15.8)</td>
<td>81 (53.8)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>592 (61.9)</td>
<td>348 (58.8)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Female</td>
<td>365 (38.1)</td>
<td>192 (52.6)</td>
<td></td>
<td>0.73 (0.56-0.95)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Primary or secondary school certificate</td>
<td>371 (38.8)</td>
<td>188 (50.7)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Ordinary national diploma (OND)</td>
<td>111 (11.6)</td>
<td>72 (64.9)</td>
<td></td>
<td>1.90 (1.22-2.98)</td>
</tr>
<tr>
<td>Higher national diploma (HND)</td>
<td>120 (12.5)</td>
<td>87 (72.5)</td>
<td></td>
<td>2.62 (1.64-4.19)</td>
</tr>
<tr>
<td>Bachelors or higher</td>
<td>317 (33.4)</td>
<td>169 (53.3)</td>
<td></td>
<td>1.12 (0.81-1.55)</td>
</tr>
<tr>
<td>Other</td>
<td>38 (4)</td>
<td>24 (63.2)</td>
<td></td>
<td>1.79 (0.89-3.60)</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>544 (56.8)</td>
<td>320 (58.8)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Town</td>
<td>323 (33.8)</td>
<td>181 (56.0)</td>
<td></td>
<td>0.88 (0.68-1.17)</td>
</tr>
<tr>
<td>Rural</td>
<td>90 (9.4)</td>
<td>39 (43.3)</td>
<td></td>
<td>0.56 (0.35-0.89)</td>
</tr>
</tbody>
</table>

Pseudo $R^2=2.79\%$.

The second column of Table 3 shows cross-tabulations between sociodemographic characteristics and vaccine uptake. There was no statistically significant difference in the COVID-19 vaccination rate by age or sex. Education was associated with vaccine uptake: respondents with an OND (n=72, 64.9% vs n=188, 50.7%; P<.001) or an HND (n=87, 72.5% vs n=188 50.7%; P<.001) were more likely to be vaccinated than respondents with a primary or secondary school certificate. Urban residence was also associated with higher vaccine uptake: respondents from rural areas were significantly less likely to have obtained the COVID-19 vaccination (n=39, 43.3% vs n=320, 58.8%; P=.02).

The third column of Table 3 shows the adjusted odds of COVID-19 vaccination. With all sociodemographic characteristics in the model, female participants were less likely to get vaccinated (aOR 0.73, 95% CI 0.56-0.95). Having an OND (aOR 1.90, 95% CI 1.22-2.98) or a HND (aOR 2.62, 95% CI 1.64-4.19) increased a respondent’s likelihood of being vaccinated. Nigerians living in rural areas were less likely to be vaccinated (aOR 0.56, 95% CI 0.35-0.89).

**Full Model: Core Motivators, Ability Barriers, and Sociodemographic Characteristics**

Table 4 shows the adjusted odds of COVID-19 vaccine uptake in Nigeria. The 3 core motivators identified earlier remained significant after adjusting for ability factors and sociodemographic characteristics. Nigerians who believed that the COVID-19 vaccination was associated with the promise of a better life were more likely to be vaccinated (aOR 3.51, 95% CI 2.23-5.52). Nigerians who felt that the vaccination would allow them to do more things they enjoyed were more likely to be vaccinated (aOR 1.97, 95% CI 1.33-2.93). Respondents’ friends’ and family members’ acceptance of their decision to get vaccinated was associated with a higher likelihood of their being vaccinated (aOR 1.62, 95% CI 1.06-2.48).
Table 4. Adjusted odds (aOR) of COVID-19 vaccination in Nigeria.

<table>
<thead>
<tr>
<th>Core motivators</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting vaccinated allows a person to live a better life (hope)</td>
<td>3.51 (2.23-5.52)</td>
</tr>
<tr>
<td>Getting vaccinated protects people from hardship (fear)</td>
<td>1.03 (0.71-1.49)</td>
</tr>
<tr>
<td>Allows people to do more things they enjoy (pleasure)</td>
<td>1.97 (1.33-2.93)</td>
</tr>
<tr>
<td>Reduces the likelihood of getting or spreading COVID-19 (pain)</td>
<td>1.17 (0.74-1.84)</td>
</tr>
<tr>
<td>Many friends and family approve vaccination (acceptance)</td>
<td>1.62 (1.06-2.48)</td>
</tr>
<tr>
<td>Most family and friends do not approve vaccination (rejection)</td>
<td>1.18 (0.83-1.67)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ability factors</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family or work responsibilities make it difficult (time)</td>
<td>0.85 (0.61-1.19)</td>
</tr>
<tr>
<td>Costs or loss of income make it difficult (money)</td>
<td>0.35 (0.24-0.50)</td>
</tr>
<tr>
<td>Physical effort makes it difficult (physical effort)</td>
<td>1.45 (1.02-2.07)</td>
</tr>
<tr>
<td>The decision to get the vaccine is difficult (mental effort)</td>
<td>1.08 (0.77-1.51)</td>
</tr>
<tr>
<td>Vaccine not available where I routinely visit (routine)</td>
<td>0.29 (0.21-0.40)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sociodemographic factors</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>1.09 (0.68-1.74)</td>
</tr>
<tr>
<td>30-39</td>
<td>1.14 (0.70-1.86)</td>
</tr>
<tr>
<td>≥40</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Female</td>
<td>0.80 (0.58-1.10)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Primary or secondary school certificate</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Ordinary national diploma (OND)</td>
<td>2.27 (1.34-3.84)</td>
</tr>
<tr>
<td>Higher national diploma (HND)</td>
<td>3.57 (2.04-6.24)</td>
</tr>
<tr>
<td>Bachelors or higher</td>
<td>1.26 (0.86-1.85)</td>
</tr>
<tr>
<td>Other</td>
<td>1.88 (0.82-4.33)</td>
</tr>
<tr>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Town</td>
<td>0.90 (0.64-1.26)</td>
</tr>
<tr>
<td>Rural</td>
<td>0.54 (0.31-0.92)</td>
</tr>
</tbody>
</table>

The relationships between ability factors and vaccine uptake remained important after adjusting for sociodemographic characteristics and core motivators. Nigerians with cost- or income-related concerns were less likely to obtain a COVID-19 vaccination (aOR 0.35, 95% CI 0.24-0.50). The lack of availability of vaccines at places they routinely visited made them less likely to get vaccinated (aOR 0.29, 95% CI 0.21-0.40).

After adjusting for motivation and ability, female participants were no longer less likely to obtain a COVID-19 vaccination. Respondents with an OND (aOR 2.27, 95% CI 1.34-3.84) or HND (aOR 3.57, 95% CI 2.04-6.24) were more likely to be vaccinated than those with primary or secondary school certificates. Rural residents were less likely to be vaccinated than residents living in cities (aOR 0.54, 95% CI 0.31-0.92).

**Discussion**

**Principal Findings**

The findings of this study show that 56.4% (n=540) of Nigerian adults who responded to the digital survey had obtained at least 1 COVID-19 vaccination by October 2022. Several core motivators were associated with vaccine uptake, after adjusting
for ability factors and sociodemographic variables. The beliefs that COVID-19 vaccination allows a person to live a better life and that it allows them to do more things that they enjoy increases the likelihood of being vaccinated. The belief that many friends and family members approve of COVID-19 vaccination is also associated with a greater likelihood of being vaccinated. Several ability barriers were also correlated with vaccine uptake, after adjusting for other variables. Respondents who felt that costs or loss of income associated with getting vaccinated made it difficult to get vaccinated were less likely to get vaccinated. The lack of availability of the COVID-19 vaccine at places respondents routinely visited was also negatively associated with vaccine uptake. The study also found a higher likelihood of Nigerians with OND or HND being vaccinated compared with those with primary or secondary school certificates and a lower likelihood of being vaccinated among rural respondents.

**Strengths and Limitations**

While this study provides useful insights for program design, its limitations should be acknowledged. The first limitation of this study is that no causal inferences can be made from it because of its cross-sectional design. This design limitation may explain an unexpected study finding: after adjusting for other variables, the physical effort required to get vaccinated was associated with a higher rate of vaccination. This finding may reflect reverse causality; those who are vaccinated may be more aware of the physical effort required to obtain a COVID-19 vaccination. Further investigation is needed to determine whether the positive relationship between the perceived physical effort required to get vaccinated and receiving a vaccination holds only for those who have been vaccinated or for the full sample.

The second limitation of this study is that variables that were not measured may be responsible for the observed relationships. For example, while the relationship between the belief that vaccination allows a person to enjoy life more and vaccine uptake is powerful, there is a possibility that unmeasured factors are driving this relationship. Thus, developing messages around how vaccination can help a person lead a fuller life and testing them through relatively inexpensive digital campaigns would be important prior to implementing an at-scale campaign that focuses on this message.

The third limitation of the study is that it is not representative of all Nigerians in the 6 states in which it was conducted. This is reflected in the higher educational status of the survey sample: about 33% (n=317) of respondents had a bachelor’s or higher education. The participants were also relatively young; about 56% (n=531) of respondents were between 18 and 29 years. Moreover, male participants comprised a higher proportion of the sample than female participants. These findings are not uncommon for digital surveys conducted in LMICs.

A strength of this study is its cost efficiency and the timeliness with which the survey was conducted compared to face-to-face household surveys. A major barrier to the use of behavioral insights by practitioners in LMICs is the cost of data collection. Behavioral research is not well-funded in LMICs. Inexpensive digital surveys could substantially increase the ability of practitioners in LMICs to use behavioral insights to develop interventions that increase vaccine uptake.

**Future Directions**

We do not know the extent to which the findings from surveys conducted by recruiting respondents through web-based advertising are comparable to the findings from population-based household surveys. Although some studies show broadly similar patterns between digital and population-based surveys [25], more research is needed to identify what types of systematic differences may exist between these 2 survey modalities. It is important, for example, to learn whether inferences from digital surveys apply to the behavior of individuals who are not on digital platforms.

Given our inability to generalize these findings beyond Nigerians who are on Facebook and Instagram, how can the findings of this survey be used? First, the findings may be used to design interventions on digital platforms as well as to evaluate the effectiveness of those interventions. Digital behavior change interventions may be evaluated by experimental studies on digital platforms that compare vaccine uptake between intervention and control groups. A growing proportion of the Nigerian population is now on Facebook and Instagram: between 31 million and 36 million Nigerians 13 and older use Facebook and Instagram each month. High exposure to messages that associate the COVID-19 vaccination with a better, more fulfilling life is achievable through advertising on digital platforms and at a fraction of the cost of advertising on traditional mass media channels such as television.

Future interventions could build upon the findings of this study by conducting qualitative research to determine which motivation or ability factors are relevant in locations where interventions are planned. A survey conducted in the most densely populated, low-income commune in Yugpognon, Cote d’Ivoire, using the FBM found broadly comparable findings: motivation and ability were powerful drivers of vaccine adoption, although the specific elements of motivation and ability that were relevant in Yugpognon were, not surprisingly, different [16].

Our findings raise several questions that should be answered through additional research. Answers to these questions may help in the design of more effective COVID-19 service delivery interventions. It would be useful, for example, to learn whether some of the places routinely visited by Nigerians are amenable to serving as COVID-19 vaccine delivery sites. Are such potential vaccine delivery sites likely to vary by age, sex, by urban or rural residence and are they suitable for cost-efficient provision of COVID-19 vaccinations?

Our sample consisted primarily of Nigerians living in cities and towns, with a minority of respondents living in rural areas. A larger proportion of rural respondents may be obtained from digital surveys that oversample rural areas. This may be done by capping the number of respondents from urban areas and allowing more time for responses to come in from rural areas. This would, however, have implications for the cost of the rural component of the survey.
Comparison to Prior Work

Overall, the findings of our study provide a very different perspective on vaccine adoption in Nigeria than what is available in the peer-reviewed literature. Recent Nigerian studies on COVID-19 vaccine uptake have primarily focused on identifying gaps in knowledge, attitudes, and beliefs associated with vaccine hesitancy [1-3]. These studies place emphasis on the perceived risk of disease as a driver of vaccine uptake, despite mixed evidence on the role of risk perception on vaccine uptake in Nigeria [5,6,9]. Several of these recent studies propose that equipping people with factually correct information will allay their concerns, increase their perceived risk of acquiring COVID-19, and lead to higher rates of vaccine adoption. By contrast, a small but rapidly growing body of work is putting vaccine adoption in a behavioral context [13-19,26]. These studies take motivation and ability for behavior change into account in explaining the range of barriers that influence immunization decisions and suggest how programs should help individuals overcome them.

Conclusions

These findings help us consider a very different approach to intervention design—one that builds upon what people want for their future, what gives them pleasure, and how they are influenced by the approval of their friends and family members. Our findings suggest that hope is more important for Nigerians than fear when it comes to vaccine adoption, social approval is more powerful than social disapproval, and enjoying life is more important than worrying about getting the disease. These findings suggest that an approach that is based on increasing the perception of hope and pleasure associated with vaccine adoption as well as increasing network members’ social approval is likely to increase COVID-19 vaccine adoption in Nigeria. Our analysis also suggests that financial considerations play an important role in the uptake of COVID-19 vaccination in Nigeria. The costs associated with reaching a vaccination site or the loss of income associated with being away from work are important determinants of vaccine adoption. Nigerians with limited flexibility at work may find it challenging to visit a vaccination site during the hours that it is open. Consistent with this finding, making COVID-19 vaccines available at places that Nigerians visit routinely may have a large impact on vaccine uptake.

The use of a behavior model to understand drivers of COVID-19 vaccine uptake in Nigeria has helped provide a different perspective on vaccine-related decision-making in Nigeria than what is currently available in the published literature. The FBM, a model of human behavior rather than a model of health behavior per se, considers a broad range of factors influencing motivation, including an individual’s hopes and fears, the sensation of pleasure or pain that they get from a particular behavior, and the social influences on them associated with their identity. The model also measures ability constraints including bandwidth-related constraints such as time or cognitive constraints, financial constraints, physical effort–related constraints, and habits or routine-related constraints. By comparison to behavior models that focus on perceived risk of and susceptibility to disease, the FBM situates behavior within the broader context of a person’s life.

Acknowledgments

The authors gratefully acknowledge the study participants for the information they provided. Funding for this study was provided by The Bill & Melinda Gates Foundation. The funders had no input on the interpretation or publication of the study results.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

IN developed and submitted the institutional review board protocol to Nigeria’s National Health Research Ethics Committee. SA and DB conceptualized the study. SA conducted the statistical analysis and wrote the first draft of the report. IN provided important context on Nigeria that enabled the interpretation of the study findings. DB developed the survey instrument and interpreted the study findings. DB and SF were responsible for acquiring funding for the study. SF and NR designed the digital data collection instruments and implemented the data collection and quality control procedures. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.

[DOCX File, 22 KB - ijmr_v13i1e47817_app1.docx ]

References


Abbreviations

- aOR: adjusted odds ratio
- FBM: Fogg Behavior Model
- HCW: health care worker
- HND: higher national diploma
- LMICs: low- and middle-income countries
- OND: ordinary national diploma

©Sohail Agha, Ifeanyi Nsofor, Drew Bernard, Sarah Francis, Nandan Rao. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.02.2024. 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.
Influence of Environmental Factors and Genome Diversity on Cumulative COVID-19 Cases in the Highland Region of China: Comparative Correlational Study

Zhuoga Deji1,2, MSc; Yuantao Tong1, PhD; Honglian Huang1,3, PhD; Zeyu Zhang1, PhD; Meng Fang4, PhD; M James C Crabbe5,6,7, PhD; Xiaoyan Zhang1, PhD; Ying Wang3, PhD

1Research Center for Translational Medicine, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China
2Information School, The University of Sheffield, Sheffield, United Kingdom
3Department of Clinical Laboratory Medicine Center, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China
4Department of Laboratory Medicine, Shanghai Eastern Hepatobiliary Surgery Hospital, Shanghai, China
5Wolfson College, Oxford University, Oxford, United Kingdom
6Institute of Biomedical and Environmental Science & Technology, University of Bedfordshire, Bedfordshire, United Kingdom
7School of Life Sciences, Shanxi University, Shanxi, China

Corresponding Author:
Ying Wang, PhD
Department of Clinical Laboratory Medicine Center
Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine
110 Ganhe Rd
Shanghai, 200437
China
Phone: 86 021 65161782
Fax: 86 021 65161782
Email: nadger_wang@139.com

Abstract

Background: The novel coronavirus SARS-CoV-2 caused the global COVID-19 pandemic. Emerging reports support lower mortality and reduced case numbers in highland areas; however, comparative studies on the cumulative impact of environmental factors and viral genetic diversity on COVID-19 infection rates have not been performed to date.

Objective: The aims of this study were to determine the difference in COVID-19 infection rates between high and low altitudes, and to explore whether the difference in the pandemic trend in the high-altitude region of China compared to that of the lowlands is influenced by environmental factors, population density, and biological mechanisms.

Methods: We examined the correlation between population density and COVID-19 cases through linear regression. A zero-shot model was applied to identify possible factors correlated to COVID-19 infection. We further analyzed the correlation of meteorological and air quality factors with infection cases using the Spearman correlation coefficient. Mixed-effects multiple linear regression was applied to evaluate the associations between selected factors and COVID-19 cases adjusting for covariates. Lastly, the relationship between environmental factors and mutation frequency was evaluated using the same correlation techniques mentioned above.

Results: Among the 24,826 confirmed COVID-19 cases reported from 40 cities in China from January 23, 2020, to July 7, 2022, 98.4% (n=24,430) were found in the lowlands. Population density was positively correlated with COVID-19 cases in all regions (ρ=0.641, P=.003). In high-altitude areas, the number of COVID-19 cases was negatively associated with temperature, sunlight hours, and UV index (P=.003, P=.001, and P=.009, respectively) and was positively associated with wind speed (ρ=0.388, P<.001), whereas no correlation was found between meteorological factors and COVID-19 cases in the lowlands. After controlling for covariates, the mixed-effects model also showed positive associations of fine particulate matter (PM2.5) and carbon monoxide (CO) with COVID-19 cases (P=.002 and P<.001, respectively). Sequence variant analysis showed lower genetic diversity among nucleotides for each SARS-CoV-2 genome (P<.001) and three open reading frames (P<.001) in high altitudes compared to 300 sequences analyzed from low altitudes. Moreover, the frequencies of 44 nonsynonymous mutations and 32 synonymous mutations were significantly different between the high- and low-altitude groups (P<.001, mutation frequency>0.1). Key nonsynonymous
mutations showed positive correlations with altitude, wind speed, and air pressure and showed negative correlations with temperature, UV index, and sunlight hours.

Conclusions: By comparison with the lowlands, the number of confirmed COVID-19 cases was substantially lower in high-altitude regions of China, and the population density, temperature, sunlight hours, UV index, wind speed, PM2.5, and CO influenced the cumulative pandemic trend in the highlands. The identified influence of environmental factors on SARS-CoV-2 sequence variants adds knowledge of the impact of altitude on COVID-19 infection, offering novel suggestions for preventive intervention.


KEYWORDS
COVID-19; environmental factors; altitude; population density; virus mutation

Introduction

Background

In recent years, the outbreak of COVID-19 has had substantial impacts on human health, social life, and economic trends worldwide. To comprehensively explore the impact of potential factors on the strength and speed of viral infection, we first used the zero-shot model to screen the literature related to broad respiratory infectious diseases, including COVID-19, to identify the major influencing factors through text mining [1]. A significant ranking list of environmental factors was obtained (Multimedia Appendix 1), in which temperature and atmospheric pressure were highly correlated with respiratory viruses. Earlier findings showed that the spread of COVID-19 was linked to various factors, including the environment, sequence variants of the virus, and government countermeasures to protect public health in the face of outbreaks [2,3].

Previous studies also investigated the correlation between COVID-19 and altitude. COVID-19 transmission in high-altitude regions appears to differ from the global pattern, with a lower number of cases reported at high altitudes [4-9]. Moreover, population density was identified as a basic factor that significantly catalyzed the spread of COVID-19 in numerous countries, including India, the United States, China, and Malaysia [10-13]. However, few studies have performed an in-depth analysis of the effects of population density, altitude, and environmental factors on the variation in the severity of the COVID-19 pandemic among regions at different altitudes.

Moreover, virus mutation is another important factor in escaping the immune protection derived from a previous infection or vaccination [14-16]. Several studies have shown that people living at high altitudes may be less susceptible to developing severe adverse effects from COVID-19, along with reduced case fatality rates [4,8]. Environmental factors have also been shown to actively influence virus mutation and to play regulatory roles in viral evolution [17-20]. Nevertheless, the underlying physiological mechanism linking virus mutation and altitude that could affect the rate of COVID-19 transmission remains unclear.

Purpose

To fill this research gap, we aimed to explore the potential factors contributing to the outcome of COVID-19 at high altitudes in China in comparison to the lowlands. Toward this end, we first assessed the contributions of altitude and population density on the total number of COVID-19 cases at different altitudes, and further explored the correlation between COVID-19 cases and environmental factors.

In addition, a more detailed correlation analysis among regions at different altitudes was performed at the city level using mixed-effects multiple linear regression models controlling for potential covariates, including meteorological and air quality factors. Furthermore, we studied the genome diversity, mutation frequency, and correlations with environmental factors in high- and low-altitude regions. Overall, we aim to provide a better understanding of the key factors that could influence the cumulative infection and transmission rate of COVID-19 in high-altitude regions of China, which can in turn help to inform establishing improved policies for preventive interventions.

Methods

Study Area

We focused on the high-altitude southwestern regions of China and several low-altitude regions from mainland China. A more detailed description of each province included in the study area along with city-level information is provided in Multimedia Appendix 2.

Collection of Confirmed COVID-19 Cases

Data on COVID-19, including total confirmed cases at the province level from January 23, 2020, to July 7, 2022, were collected from the Dingxiangyuan (DXY-DX Doctor) website [21]; historical cases at the city level over the same study period were collected using the R package “nCovid2019.” The population density for each region was calculated by the following equation:

\[
\text{Population density} = \frac{\text{population size}}{\text{area of the land}} \quad (1)
\]

The normalized daily number of confirmed COVID-19 cases was calculated using the following equation:

\[
\text{Normalized COVID-19 cases} = \frac{\text{total number of confirmed cases}}{\text{population density}} \quad (2)
\]

To reflect the COVID-19 infection situation in high-altitude regions of China during the study period, descriptive statistics were compiled for the daily average meteorological and air quality parameters, as shown in Multimedia Appendix 3.
Collection of Population, Meteorological, and Air Quality Factors

To analyze the correlation between altitude and COVID-19 cases in China, altitude information of highland regions (>1500 m) and lowland regions (<1500 m) was collected from the topographic map [22]. Meteorological factors were collected from the World Weather Online site [23]. Air quality variables were collected from the Statistical Yearbook for each municipality or province, along with associated data on population and land areas. Multimedia Appendix 4 provides a more detailed description of all these variables.

Collection of SARS-CoV-2 Genome Data

SARS-CoV-2 whole-genome sequencing data were collected from the GISAID (Global Initiative on Sharing All Influenza Data) website [24], including 300 sequences from high-altitude regions and 300 sequences randomly selected over the same time period from low-altitude regions. The accession numbers of the sequences included in this analysis are listed in Multimedia Appendix 5.

Statistical Analysis

The Kolmogorov-Smirnov test was performed for each variable (see Multimedia Appendix 6) with the null hypothesis of a normal distribution; since the P value of each variable was less than .05, the null hypothesis was rejected, indicating that the test distribution was not normally distributed. Therefore, we used the nonparametric Spearman correlation coefficient for the correlation analysis of environmental factors and mutation frequency with COVID-19 cases using SPSS software.

After taking into account the correlations for each of the independent variables, a mixed-effects multiple linear regression was used for an adjusted correlation analysis, with the model defined as follows:

\[ Y = X\beta + Z\mu + \epsilon \]  

where \( X\beta \) represents the fixed-effects set in this study, including the covariates meteorological and air quality, and \( Z\mu \) represents an \( N \times M \) design matrix containing each individual group (N) for each covariate (M) of the random effects. Four sets of mixed-effects models were analyzed using R software (version 4.2.1). Separate models were first run for each of the 12 random effects (environmental covariates), followed by two separate models that included all 6 meteorological indicators or air quality indicators as fixed effects simultaneously. The final model included all 12 indicators simultaneously.

Genome Analysis

The nucleotide sequences of the whole genome of SARS-CoV-2 were aligned to the reference sequence Wuhan-1 (NC_045512.2) using minimap2 2.1.7-r974. All mapped sequences were merged back with all others in a single alignment bam file. Variant calling was performed using bcftools mpileup v1.91. Gene sequences of SARS-CoV-2 were extracted and translated into amino acid sequences, which were aligned to the reference sequence by ClustalW. Variant calling was computed using an in-house-developed R script.

Sequence Diversity Calculation

Sequence diversity was calculated using the Shannon entropy (\( S_n \)) index in R software (version 4.2.1), which measures the diversity of nucleotides, amino acids, and their respective variant frequencies. The diversity of each nucleotide position (nucleotide 1 to 29,903) was calculated as the \( S_n \) according to the following formula [25]:

\[ S_n = \sum (p_i \log p_i) / \log N \]  

where \( p_i \) represents the relative frequency of nucleotides or deletion at this position and \( N \) represents the total number of sequences. All high-altitude samples were compared with the low-altitude samples to identify differential sequence variations at both the nucleotide and amino acid levels.

Ethical Considerations

This study is based on an analysis of existing data and therefore received exemption from ethical approval (reference number 041797). Some data sets were slightly modified after the study received exemption from ethical approval; however, the updated data were collected from the same publicly available database. According to the University Research Ethics Policy, “the project does not involve human participants which will only use publicly available anonymized data; a project which will only use existing clinical or research data that has been robustly anonymized such that it no longer constitutes personal data” (pp. 3.1.10, p. 20-21); therefore, ethical approval was not required for the above circumstance. In addition, Article 32 of the Ethical Review Measures for Human Life Science and Medical Research of the People’s Republic of China declares that “research involving human life science and medicine under the circumstances of using anonymized information data, legally obtained publicly available data, or data generated through observation without interference with public behavior, which do not cause harm to the human body and do not involve sensitive personal information or commercial interests, can be exempt from ethical review to reduce unnecessary burdens on researchers and promote the development of human life science and medical research” (chapter 3). Therefore, this study falls within the exempt category from the institutional review board. Further ethical approval by a review committee was not required since the data utilized in this study have been sourced from a publicly accessible database, ensuring full anonymization, and the research process involved no direct interaction with human subjects, solely relying on the analysis of pre-existing and publicly available data, as mentioned in the above legislation.

Results

Correlation Between Altitude and COVID-19 Cases

During the study period from January 23, 2020, to June 7, 2022, a total of 732 COVID-19 cases were officially reported in the high-altitude region of China, which excluded one imported case in Tibet. The total numbers of confirmed COVID-19 cases were lower in the high-altitude regions, including Tibet, Qinghai, and Gansu, than in the low-altitude regions (Figure 1A); despite variation among regions, 98.4% (24,430/24,826) of the total COVID-19 cases were found in the lowland regions (Figure
Taking into account the hidden factors of the dimension of population size (Multimedia Appendix 2) in different areas, a significant positive correlation between confirmed COVID-19 cases and population density (Figure 1C) was found during the study period.

Figure 1. Pattern of confirmed COVID-19 cases at high altitudes within the study period. (A) Geographic patterns of confirmed COVID-19 cases in China as of July 7, 2022. (B) Comparison of confirmed COVID-19 cases reported in different cities at different altitudes. (C) Linear regression analysis between population density and confirmed COVID-19 cases in China.

Correlation Between Meteorological Factors and Confirmed COVID-19 Cases in High- and Low-Altitude Regions

To gain a better understanding of the correlation of meteorological factors with the spread of COVID-19 in high-altitude regions when compared with that in the lowlands, we performed a comparative correlation analysis of each variable (Figure 2); the detailed numeric results are provided in Multimedia Appendix 7. In the highlands of China, average temperature ($P=.003$), sunlight hours ($P<.001$), and UV index ($P=.009$) were negatively correlated with the number of confirmed COVID-19 cases, whereas the wind speed ($r=0.388$, $P<.001$) was positively correlated with confirmed cases.
Figure 2. Spearman correlation analysis of confirmed COVID-19 cases and meteorological factors in high-altitude and low-altitude regions in China. The color represents each region group. Area charts represent frequencies and the scatterplots with lines represent the correlation. \( P \) values are based on 2-tailed tests; also see Multimedia Appendix 7.

Correlation Between Air Quality Factors and Confirmed COVID-19 Cases in High- and Low-Altitude Regions

As shown in Figure 3, we found significant correlations between air quality factors and COVID-19 cases at high and low altitudes. In the highlands, COVID-19 cases were positively correlated with fine particulate matter (PM2.5; \( P = .002 \)), coarse particulate matter (PM10; \( P < .001 \)), nitrogen dioxide (NO2; \( P = .02 \)), and ozone (O3; \( P < .001 \)). In the lowlands, a negative correlation was found between confirmed COVID-19 cases and air quality index (AQI; \( P = .01 \)), whereas positive correlations were found for PM2.5 (\( P < .001 \)), NO2 (\( P < .001 \)), and carbon monoxide (CO; \( P < .001 \)); in contrast to the pattern in the highlands, the average daily concentrations of PM10 and O3 were not significantly correlated with COVID-19 cases in lowland regions of China. The detailed results are shown in Multimedia Appendix 7.
Correlations of Meteorological and Air Quality Factors With Normalized Confirmed COVID-19 Cases After Covariate Adjustment

The results of the mixed-effects multiple linear regression among the 12 provinces (Multimedia Appendix 2) at different altitudes are summarized in Table 1. In model 1, including each of the fixed-effect indicators in 12 separate models, a lower number of confirmed COVID-19 cases was significantly associated with average temperature, sunlight hours, UV index, air pressure, and air quality factors, including PM2.5, O3, and CO. However, in model 2, including all of the meteorological covariates considered simultaneously, only average temperature, UV index, and air pressure were significantly associated with the normalized confirmed COVID-19 cases. In model 3, including all of the air quality covariates considered simultaneously, only the average concentrations of O3 and CO were significantly correlated with confirmed COVID-19 cases. Finally, in model 4, including all 12 environmental indicators considered simultaneously, only average temperature, wind speed, and CO were significantly associated with confirmed COVID-19 cases.
### Table 1. Mixed-effects multiple linear regression of the associations of normalized COVID-19 cases with meteorological and air quality factors including adjustment for covariates.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Model 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Model 2&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Model 3&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Model 4&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>P value</td>
<td>Estimate (SE)</td>
<td>P value</td>
</tr>
<tr>
<td>Average temperature (°C)</td>
<td>-15.41 (2.55)</td>
<td>.001</td>
<td>-21.44 (7.47)</td>
<td>.004</td>
</tr>
<tr>
<td>Sunlight hours</td>
<td>-45.31 (9.50)</td>
<td>.001</td>
<td>-12.55 (15.04)</td>
<td>.40</td>
</tr>
<tr>
<td>UV index</td>
<td>-201.24 (27.53)</td>
<td>.001</td>
<td>-257.41 (71.06)</td>
<td>.001</td>
</tr>
<tr>
<td>Wind speed</td>
<td>7.934 (10.81)</td>
<td>.46</td>
<td>16.72 (11.27)</td>
<td>.13</td>
</tr>
<tr>
<td>Humidity</td>
<td>-0.071 (1.42)</td>
<td>.96</td>
<td>-1.99 (1.81)</td>
<td>.27</td>
</tr>
<tr>
<td>Pressure</td>
<td>17.64 (2.51)</td>
<td>.001</td>
<td>20.35 (5.51)</td>
<td>.001</td>
</tr>
<tr>
<td>AQI&lt;sup&gt;f&lt;/sup&gt;</td>
<td>-1.45 (0.67)</td>
<td>.03</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>PM2.5&lt;sup&gt;g&lt;/sup&gt;</td>
<td>4.71 (0.99)</td>
<td>.001</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>PM10&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.36 (0.47)</td>
<td>.43</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>SO2&lt;sup&gt;i&lt;/sup&gt;</td>
<td>6.93 (6.01)</td>
<td>.24</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>NO2&lt;sup&gt;j&lt;/sup&gt;</td>
<td>1.88 (1.612)</td>
<td>.24</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>O3&lt;sup&gt;k&lt;/sup&gt;</td>
<td>2.29 (0.77)</td>
<td>.002</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>CO&lt;sup&lt;l&lt;/sup&gt;</td>
<td>375.4 (75.5)</td>
<td>.001</td>
<td>_</td>
<td>_</td>
</tr>
</tbody>
</table>

<sup>a</sup>Model 1 includes each of the fixed effects run in 12 separate models.

<sup>b</sup>Model 2 includes all meteorological covariates considered simultaneously.

<sup>c</sup>Model 3 includes all air quality covariates considered simultaneously.

<sup>d</sup>Model 4 includes all 12 environmental indicators considered simultaneously.

<sup>e</sup>Excluded from model.

<sup>f</sup>AQI: air quality index.

<sup>g</sup>PM2.5: fine particulate matter.

<sup>h</sup>PM10: coarse particulate matter.

<sup>i</sup>SO2: sulfur dioxide.

<sup>j</sup>NO2: nitrogen dioxide.

<sup>k</sup>O3: ozone.

<sup>l</sup>CO: carbon monoxide.

### Diversity of SARS-CoV-2 Sequences in High- and Low-Altitude Groups

Comparable sequence diversity was found among the full-length SARS-CoV-2 genomes and 12 genes of SARS-CoV-2 in the high- and low-altitude groups (Figure 4). In the high-altitude group, the Sn values of nucleotides for each site were significantly lower than those in the low-altitude group (Figure 4A; *P*<.001). Likewise, the Sn values of amino acids for three open reading frames (ORFs; ORF1b, ORF7a, and ORF7b) in the high-altitude group were lower than those in the low-altitude group (Figure 4B; *P*<.001). Compared to those of the low-altitude group, the Sn values of the S and N genes were higher in the high-altitude group.
Figure 4. Mean sequence diversity across high-altitude (red) and low-altitude (blue) groups identified by Shannon entropy in nucleotides and different amino acids of the SARS-CoV-2 genome. (A) Mean sequence diversity identified by Shannon entropy in nucleotides of the whole SARS-CoV-2 genome. (B) Mean sequence diversity identified by Shannon entropy in different amino acids. Error bars indicate SEM. *P values are based on the Wilcoxon signed rank test.

Comparison of SARS-CoV-2 Gene Variant Frequencies in the High- and Low-Altitude Groups

To determine the gene variant frequency difference between the high- and low-altitude groups, the Wilcoxon signed rank test was used to evaluate the significance of variant frequency in each of the ORFs (Figure 5A). The amino acid mutation frequencies for the three ORFs (ORF1b, ORF7a, and ORF7b) were significantly lower in the high-altitude group than in the low-altitude group (*P<.001) but were higher for the S and N genes (*P<.001). Among these, 44 nonsynonymous mutations and 32 synonymous mutations were found between the high- and low-altitude groups (Figure 5B; "P<.001; mutation frequency>0.1). There was a greater proportion of nonsynonymous mutations in the low-altitude group than in the high-altitude group, whereas the opposite pattern was found for synonymous mutations. The –log10 P values for the differences in nonsynonymous and synonymous mutations for each ORF between the high- and low-altitude groups are shown in Figure 5C. Compared to the result in Figure 5A, the synonymous mutations of the S genes and N genes had relatively higher significance scores in the high-altitude group.

Figure 5. Distribution of mutations of SARS-CoV-2 in the high- and low-altitude groups. (A) Mean mutation frequency of open reading frames (ORFs) in the high- and low-altitude groups. (B) Comparison of frequencies of nonsynonymous and synonymous mutations between the high- and low-altitude groups. (C) Significance scores of nonsynonymous and synonymous mutations in each ORF. ***P<.001.
Correlation Analysis Between Mutation Frequency and Environmental Factors

The impact of 7 environmental factors on the nonsynonymous mutation frequency was analyzed for all sequences (Figure 6 and Multimedia Appendix 8). Most of the nonsynonymous mutations had a positive correlation with altitude, wind speed, and atmospheric pressure, but had negative correlations with UV index, relative humidity, and sunlight hours. The highest correlation between mutation frequency and environmental factors was found for altitude with N_G29427A, ORF1b_C20320T, and ORF1a_C1191T ($\rho=0.53$, $P<.001$); followed by UV index with ORF8_C28093T and S_A23063T ($\rho=0.42$, $P<.001$); and temperature with ORF8_C28093T, S_C23604A, S_A23063T, and ORF1a_G6671A. Some significant nonsynonymous mutational events were also discovered in this study, including C22323T in the receptor-binding domain (RBD) of the S gene, which had a positive correlation with altitude ($\rho=0.32$, $P<.001$), indicating higher frequencies at high altitude. S gene mutations A23063T in the RBD region and C23604A in the fusion peptide (FP) region had higher frequencies at low altitude, and demonstrated a significant positive correlation with atmospheric pressure ($\rho=0.34$ and $\rho=0.31$, respectively; both $P<.001$) and a significant negative correlation with temperature and UV index ($\rho=-0.42$ and $\rho=-0.41$, respectively; both $P<.001$). Furthermore, co-occurring point mutations in the N gene, specifically G28881A, G28882A, and G28883C (R203K and G204R), had a higher mutation frequency at low altitude and were positively related to altitude ($\rho=0.43$, $P<.001$).

Figure 6. Spearman correlation coefficients of environmental factors and mutation frequencies among different sites in the high- and low-altitude groups. The color represents the strength of the Spearman correlation. Also see Multimedia Appendix 8.

Discussion

Principal Findings

This study found that the lower number of confirmed COVID-19 cases in high-altitude regions of China may be related to population density and environmental factors. By further exploring SARS-CoV-2 sequences, we found different mutation frequencies in the high- and low-altitude regions, which were also correlated with environmental factors.

Population density is one of the most effective predictors related to the regional pandemic; a larger population greatly increases the infection and transmission rates of COVID-19 [28]. Therefore, our results suggest that when analyzing the correlation between COVID-19 cases in regions with different population dimensions and other possible factors, it is important to take into account the normalized number of COVID-19 cases to study the pandemic trend.

Meteorological factors may also influence COVID-19 transmission and cumulative infection in high-altitude regions. Previous studies indicated that higher temperature and UV radiation could contribute to a decrease in new cases of COVID-19 infection at high altitude [29-31]. This is supported by our Spearman correlation analysis showing that the average daily temperature, sunlight hours, and UV index were negatively related to the normalized COVID-19 cases in high-altitude regions. A high-altitude environment is distinguished by lower temperatures compared with those of low-altitude regions, along with significant differences between daylight hours, air dryness, and levels of UV light radiation [32-34]. Importantly, our study further found that the wind speed in the high-altitude region of China was significantly associated with the reducing trend of COVID-19 cases, which is consistent with earlier studies conducted in Italy, New York, and Singapore [2,35]. Analyzing the correlation between covariates in regions at different altitudes with different population dimensions showed positive correlations of average temperature and wind speed with normalized confirmed COVID-19 cases, in line with previous research.

Air pollution is widely recognized as a major risk factor for respiratory infection in humans, which has also played a significant role in the spread of COVID-19. Previous research suggested that the concentrations of NO2, PM2.5, PM10, and O3 are positively correlated with the number of confirmed COVID-19 cases [33,34,36-38]. Our Spearman correlation results align with these previous findings. NO2 is often linked...
with vehicle emissions and energy production [39], which is also an irritant of human respiratory diseases. Early research findings [40] support the results of this study, suggesting that travel restrictions should be among the specific actions implemented to reduce the spread of COVID-19. Previous studies indicated that PM2.5 and PM10 levels were positively correlated to the number of new daily confirmed COVID-19 cases in mainland China [41-43]. Consistently, our findings showed that lower average concentrations of PM2.5 and PM10 were related to a lower number of confirmed COVID-19 cases in the highlands.

It is clear that all of the above correlated factors may have influenced the cumulative infection trends of the pandemic in the highlands of China. The main sources of CO are motor vehicles and industrial source emissions; consequently, CO is closely related to population density and human activity. Therefore, taking into account the mixed correlations between all the covariates of focus in this study by including the related factors into a single mixed-effects model, among the air quality factors, only the average concentration of CO was positively associated with the normalized number of confirmed COVID-19 cases during the study period. This result also aligns with previous research conducted in China [43,44].

In terms of variations in the SARS-CoV-2 genome at high and low altitudes, the whole-genome sequences at high altitude showed lower diversity based on the mean Sn values of nucleotides and different amino acids compared to those in the low-altitude sites. Previous studies have suggested an impact of high altitude on the pathophysiology of COVID-19 [44,45], implying improved tolerance to SARS-CoV-2 infection for residents of high-altitude regions. Our study identified differences in the frequencies of SARS-CoV-2 sequence variants in various ORFs between high- and low-altitude regions, along with correlations between mutation frequency and environmental factors. Most of the nonsynonymous mutations identified have been reported previously and have important biological implications. We found that only the C22323T variant of the S gene had a positive correlation with altitude, with a greater frequency in the high-altitude group. Sivasubramanian et al [46] found that the S2255F (C22323T) variant could reduce the affinity between the S protein and antibodies. We also identified the variant A23063T in the RBD region and the variant C23604A in the FP region of the S gene, with higher mutation frequencies detected at low altitude, showing a significantly positive correlation with atmospheric pressure and a strongly negative correlation with temperature and UV index. Previous studies found that the N501Y (A23063T) and Q493H variants enhanced the binding affinity to the human angiotensin converting-enzyme 2 receptor, thereby increasing infectivity [47], and that the Omicron peptide with variants N679K and P681H (C23604A) might increase viral infectivity and transmissibility [48]. These results suggest that atmospheric pressure, UV index, and temperature may affect the infectivity of COVID-19 driven by specific mutations of the S gene in low-altitude regions. In addition, three co-occurring variants (G28881A, G28882A, G28883C) in the N gene had higher mutation frequencies in the low-altitude group and were positively related to altitude; these mutations were previously reported to destabilize and decrease the overall structural flexibility of the SARS-CoV-2 genome [48].

Notably, the potential impact of the strategy implemented by the government in response to the pandemic in high-altitude regions on the observed differences should not be ignored. This study only included data collected from January 22, 2020, to March 19, 2021, during which time there was only one imported COVID-19 case in Tibet that was excluded from the analysis. However, after several consecutive months of no new cases, a newly confirmed case was reported in Tibet on August 7, 2022. Since then, eight new cases have been added with a total of 1437 confirmed cases reported to date [21]. Several reasons for this new outbreak are worth discussing. First, the virus underlying the epidemic in Tibet is Omicron BA.2.76, which has characteristics of strong, fast transmission ability and the potential to more readily escape from immune protection. Second, the first discovered infection on August 7 was a close contact of a family, with suspected transmission occurring at a family gathering, and August is the peak tourist season in Tibet. Thus, under the conditions of relatively relaxed travel restrictions, the close contact and population flow greatly increased the spread of the virus. Third, the lack of epidemic prevention and medical appliances in high-altitude regions limited access to health care systems, and the limited capacity for viral testing and contact tracing worsened the situation of the pandemic in Tibet. Therefore, effective and accurate government policies have played an important role in preventing Tibet from being affected by the epidemic, and the government policy in response to the pandemic must be considered; indeed, the response needs to be strengthened, although China indicated that they had ceased counting COVID-19 cases on December 23, 2022, and stated that the COVID-19 pandemic is effectively now over [49].

Limitations

Our study has some limitations. First, among the data available for confirmed COVID-19 cases, we only excluded the one known imported case in Tibet during the study period as of March 19, 2021, whereas the collected COVID-19 data from the other four areas may include both local and imported cases. This is mainly due to a lack of accurate publicly available COVID-19 data at the city level. Similarly, the incompleteness of some city-level information of genome sequences also led to difficulties in studying the differences in genetic variants among cities in the same province. Second, we aimed to evaluate whether virus mutations are associated with particular viral lineages. The sequences of SARS-CoV-2 in the high- and low-altitude groups represented 33 different virus lineages (Multimedia Appendix 9): B.1.617.2, AY.122, and B.1.36.16 were the major lineages, and the mutations of each lineage in which the mutation frequency was greater than 0.5 are listed in Multimedia Appendix 10. The number of different lineages between high- and low-altitude groups is also unbalanced. Third, we lacked the clinical information to study the potential relationships between mutations and clinical outcomes of infection at high and low altitudes. Therefore, future studies should focus on exploring the underlying mechanisms contributing to the links between patterns of SARS-CoV-2 mutation and case numbers at different altitudes.
Conclusions

Compared to that in the lowland area of China, the total number of confirmed COVID-19 cases in the highland was substantially lower. Population density and environmental factors, including average temperature, sunlight hours, UV index, wind speed, NO2, O3, PM2.5, and CO, were identified as indicators with a significant influence on the cumulative pandemic trend in the highlands. Among these factors, average temperature and CO were identified as the major meteorological and air quality factors associated with the spread of COVID-19 infection in China. Furthermore, we identified different mutations from SARS-CoV-2 isolates between high- and low-altitude regions, and there was a significant impact of environmental factors on virus mutation. Overall, this study adds important knowledge of the impact of altitude and related environmental factors on the cumulative infection rate of COVID-19, providing novel suggestions for preventive interventions.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81972914, 81573023), Innovation Group Project of Shanghai Municipal Health Commission (2019CXJQ03), Fundamental Research Funds for the Central Universities (22120200014), and Shanghai “Rising Stars of Medical Talent” Youth Development Program (2019-72).

Data Availability

The data sets generated and analyzed during the study are provided in the Multimedia Appendix files.

Authors’ Contributions

YW and XZ supervised the study and reviewed the draft of the paper. ZD and YT performed the data collection, main statistical analysis, and manuscript writing. HH, MF, and MJCC participated in manuscript verification, revisions, and editing. All authors had full access to all the data in the study and accept the responsibility to submit it for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Key environmental factors associated with respiratory infectious diseases reported in the literature.

[DOCX File, 52 KB - ijmr_v13i1e43585_app1.docx ]

Multimedia Appendix 2

Detailed characteristics of the regions included in the study area in the high- and low-altitude groups.

[DOCX File, 19 KB - ijmr_v13i1e43585_app2.docx ]

Multimedia Appendix 3

Descriptive statistics of all variables collected during the study period.

[DOCX File, 14 KB - ijmr_v13i1e43585_app3.docx ]

Multimedia Appendix 4

Description of all variables analyzed in this study.

[DOCX File, 14 KB - ijmr_v13i1e43585_app4.docx ]

Multimedia Appendix 5

SARS-CoV-2 sequences downloaded from the GISAID database for analysis in this study.

[XLSX File (Microsoft Excel File), 86 KB - ijmr_v13i1e43585_app5.xlsx ]

Multimedia Appendix 6

One-sample Kolmogorov-Smirnov test of normality for each variable.

[XLSX File (Microsoft Excel File), 11 KB - ijmr_v13i1e43585_app6.xlsx ]

Multimedia Appendix 7

Spearman correlation analysis of meteorological factors and air quality factors with confirmed COVID-19 cases in high-altitude and low-altitude regions.

[XLSX File (Microsoft Excel File), 33 KB - ijmr_v13i1e43585_app7.xlsx ]
Spearman correlation analysis of environmental factors and mutation frequency in the high- and low-altitude groups.

Multimedia Appendix 9
Frequencies of nonsynonymous and synonymous mutations of the main lineages of SARS-CoV-2 in the high- and low-altitude groups.

Multimedia Appendix 10
Number of samples of different SARS-CoV-2 lineages in the high-altitude (red) and low-altitude (blue) regions.

References


26. Ethic policy governing research involving human participants personal data and human tissue: general principles and statements. The University of Sheffield. URL: https://www.sheffield.ac.uk/rpi/ethics-integrity/policy [accessed 2024-03-13]

27. Ethical review measures for human life science and medical research, Chapter 3, Article 32. People's Republic of China.


Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQI</td>
<td>air quality index</td>
</tr>
<tr>
<td>CO</td>
<td>carbon monoxide</td>
</tr>
<tr>
<td>FP</td>
<td>fusion peptide</td>
</tr>
<tr>
<td>GISAID</td>
<td>Global Initiative on Sharing All Influenza Data</td>
</tr>
<tr>
<td>NO2</td>
<td>nitrogen dioxide</td>
</tr>
<tr>
<td>O3</td>
<td>ozone</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PM2.5</td>
<td>fine particulate matter</td>
</tr>
<tr>
<td>PM10</td>
<td>coarse particulate matter</td>
</tr>
<tr>
<td>RBD</td>
<td>receptor-binding domain</td>
</tr>
<tr>
<td>Sn</td>
<td>Shannon entropy</td>
</tr>
</tbody>
</table>

©Zhuoga Deji, Yuantao Tong, Honglian Huang, Zeyu Zhang, Meng Fang, M James Crabbe, Xiaoyan Zhang, Ying Wang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 25.03.2024. 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.
The Association Between Depressive Symptoms and the Weekly Duration of Physical Activity Subset by Intensity and Domain: Population-Based, Cross-Sectional Analysis of the National Health and Nutrition Examination Survey From 2007 to 2018

Josheil K Boparai¹, BSc; Sarah Dunnett¹, BA; Michelle Wu¹, BSc, MSc; Vanessa K Tassone¹, BA; Sophie F Duffy¹, BSc, MSc; Valentina Zuluaga Cuartas¹, BSc; Ziming Chen¹, BSc, MSc; Hyejung Jung², BSc, MSc; Catherine M Sabiston³, BSc, MHK, PhD; Wendy Lou², PhD; Venkat Bhat¹,4,5,6, MSc, MD

¹Interventional Psychiatry Program, St. Michael’s Hospital, Toronto, ON, Canada
²Department of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
³Department of Kinesiology, Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, ON, Canada
⁴Mental Health and Addictions Services, St. Michael’s Hospital, Toronto, ON, Canada
⁵Department of Psychiatry, University of Toronto, Toronto, ON, Canada
⁶Institute of Medical Science, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Venkat Bhat, MSc, MD
Interventional Psychiatry Program
St. Michael’s Hospital
193 Yonge Street
Floor Number 6-013
Toronto, ON, M5B1M8
Canada
Phone: 1 4163604000 ext 76404
Email: venkat.bhat@utoronto.ca

Abstract

Background: Prior literature suggests a dose-response relationship between physical activity (PA) and depressive symptoms. The intensity and domain of PA are suggested to be critical to its protective effect against depression; however, existing literature has shown mixed results.

Objective: The purpose of this population-based study is to examine the associations between depressive symptoms and weekly duration of (1) total PA and (2) PA subset by intensity, domain, or both.

Methods: A cross-sectional analysis of National Health and Nutrition Examination Survey data from 2007 to 2018 was conducted using multivariable logistic and linear regression models and survey weights. Participants (N=29,730) were 20 years and older and completed the Physical Activity Questionnaire and Depression Screener. The primary outcome was the presence of depressive symptoms, and the secondary outcomes were cognitive-affective and somatic symptoms of depression.

Results: Participants (N=29,730) had a weighted mean age of 47.62 (SD 16.99) years, and 15,133 (51.34%) were female. On average, participants without depressive symptoms engaged in 10.87 hours of total PA per week, whereas participants with depressive symptoms engaged in 8.82 hours (P<.001). No significant associations were seen between the weekly duration of total PA and depressive symptom odds, somatic, or cognitive-affective symptoms (all P>.05). Participants with an increased weekly duration of recreational PA had decreases in depressive symptom odds (adjusted odds ratio [aOR] 0.965, 95% CI 0.944-0.986) and in somatic (aβ=–0.016, 95% CI –0.022 to –0.009) and cognitive-affective (aβ=–0.015, 95% CI –0.023 to –0.007) symptoms. When recreational PA was subset by intensity, participants with an increased weekly duration of vigorous-intensity recreational PA had decreases in depressive symptom odds (aOR 0.926, 95% CI 0.883-0.972) and in somatic (aβ=–0.021, 95% CI –0.032 to –0.010) and cognitive-affective (aβ=–0.022, 95% CI –0.035 to –0.009) symptoms. However, significant associations were not seen for the weekly duration of work-related, moderate- or vigorous-intensity PAs (all P>.05).
Conclusions: Findings suggest that recreational, not work-related PA is associated with reduced symptoms of depression. Future studies should explore the impact of the different types and contexts of PA on depressive symptomatology.

(KEYWORDS)

depressive disorder; exercise; physical activity intensity; recreational physical activity; work-related physical activity; National Health and Nutrition Examination Survey; NHANES; nutrition surveys; recreational activity; physical activity; depression

Introduction

Depression is the leading cause of disability, affecting 280 million individuals globally, with limited treatment accessibility due to stigmatization and financial barriers [1]. Physical activity (PA) can have a positive impact on mood and mental health [2] among individuals with a clinical diagnosis of depression, as well as nonclinical community subsamples [3], providing the potential for more accessible and cost-effective therapeutic modalities that improve well-being [4,5]. Further, PA can be protective against depression. A study on a community sample from Alameda County in the United States showed that individuals with low baseline levels of PA had a significantly higher risk of depression at follow-up compared to those with high baseline levels of PA [6].

PA intensity is determined by energy expenditure, which is expressed as multiples of the metabolic equivalent of task, and can be subdivided into 3 categories: light, moderate, or vigorous [7]. There is substantial variability in the literature regarding the optimal intensity for reducing depressive symptoms. Light-intensity PA has been reported to reduce depression as effectively as moderate-intensity PA among older adults [8]. Another study conducted with college students aged 15-24 years found an association between light-intensity PA and reduced depressive symptoms but not moderate- or vigorous-intensity PA [9]. A recent review suggested that moderate-intensity PA has a greater protective effect on depression than higher-intensity PA [10]. Despite these inconsistencies, the Centers for Disease Control and Prevention (CDC) recommends 150 minutes of moderate-intensity or 75 minutes of vigorous-intensity exercise per week or a combination of the 2 intensities to gain health benefits, including a reduced depression risk [7]. In line with these recommendations, a recent systematic review and meta-analysis found an inverse curvilinear association between PA and incident depression, wherein individuals with an activity volume of 150 or 75 minutes per week of moderate-intensity exercise reported a 25% and 18% lower risk of depression, respectively [11]. Similarly, one study using survey data of respondents to the Scottish Health Survey reported that 20 minutes per week of any type of PA (eg, low-intensity walking) is associated with a lower risk of psychological distress [12], while another study using data from the Swedish National March Cohort found that replacing 30 minutes of sedentary behavior per day with 30 minutes of light-intensity or moderate- to vigorous-intensity PA reduced the odds of depression by 13% and 19%, respectively [13]. This suggests that the benefits of PA may be noticeable well below the CDC’s recommended levels.

PA can also be classified by domain, wherein it can be recreational (leisure) or work-related (nonleisure). PA domain has a significant impact on a person's psychosocial experience [14,15]. Recreational PA is considered more enjoyable, variable, and autonomous than work-related PA, which is often "obligatory, repetitive, or routine" [14]. Recreational PA also impacts one's perceived level of control and can act as a distraction from negative preoccupations [16]. As a result, studies have found leisure-time PA to be associated with a lower prevalence of depressive symptoms compared to nonleisure PA [14,17], suggesting that work-related PA may act as a source of stress [18]. Additionally, reducing sedentary work while controlling for leisure-time PA [19] and engaging in a greater frequency of domestic activities (eg, housework and gardening) have both been associated with lower odds of psychological distress [12]. There are also studies that have found no association between work-related PA and depression [15,20]. Given these conflicting results, it is essential to further explore the association between PA domain and depressive symptomatology.

Depression is a heterogeneous disorder characterized by several phenotypic manifestations. One modality of subtyping is based on cognitive-affective (eg, negative mood) and somatic symptom (eg, fatigue) domains [21,22]. By subdividing depression into the 2 symptom domains, we can better understand the mechanisms by which PA is associated with depression and can ultimately inform public health guidelines.

This study explored the relationship between PA intensity and domain and depressive symptoms using the National Health and Nutrition Examination Survey (NHANES). We investigated three primary aims: how are depressive symptoms related to the weekly duration of (1) total PA; (2) PA, subset by intensity; and (3) PA, subset by domain? We hypothesized that participants with a higher weekly duration of total PA, moderate- and vigorous-intensity PA, or recreational PA will experience a decrease in depressive symptom odds. Our secondary aim was to investigate the association between depressive symptoms and PA, subset by intensity and domain. We hypothesized that participants with a higher weekly duration of moderate- and vigorous-intensity recreational PA, but not moderate- and vigorous-intensity work-related PA, will experience lower depressive symptom odds. In an exploratory aim, we investigated the relationship between depressive symptom subgroups and the weekly duration of total PA and PA subset by intensity, domain, or both.
Methods

Study Population
The study data were obtained from the 2007-2018 NHANES, a cross-sectional survey administered by the National Center for Health Statistics (NCHS) and the CDC [23]. NHANES assesses the health and nutritional status of US civilians (excluding institutionalized individuals) via a home interview and a health examination. Sample selection included counties or groups of neighboring counties as primary sampling units, followed by selection of segments within primary sampling units, selection of households within segments, and finally selection of individuals within households [23]. This study population consisted of respondents 20 years and older who completed the Physical Activity Questionnaire (PAQ) and the Mental Health—Depression Screener, which uses a standardized depression scale (ie, the Patient Health Questionnaire-9 [PHQ-9]).

Ethical Considerations
The NHANES protocol was reviewed by the NCHS Research Ethics Review Board (protocols 2005-06, 2011-17, and 2018-01) with all participants providing informed consent prior to the interview and examination. The original consent provided by NHANES participants includes the use of their data for secondary analyses. As such, the secondary use of these data does not require additional consent from participants. The NHANES data are deidentified with the omission of all direct identifiers and characteristics that might lead to identification. This analysis did not receive approval from an institutional review board since the Data User Agreement provided by the NCHS specifies that the data in the data set can be used for statistical reporting and analysis [24].

Exposure
Exposure variables in this study included the weekly duration of total PA and weekly duration of PA subset by intensity (moderate and vigorous), domain (recreational and work-related), or both. Each PA variable was analyzed on a continuous scale of total hours per week. For details on the questionnaire items used for the exposure variables, see Table S1 in Multimedia Appendix 1.

Primary Outcome Measure
The primary outcome measure for this study was the sum of items 1 to 9 of the PHQ-9 from the Mental Health—Depression Screener. The PHQ-9 assessed the frequency of depressive symptoms over the past 2 weeks based on the Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition major depressive disorder diagnostic criteria [25]. Responses were given on a 4-point Likert scale ranging from 0=not at all to 3=nearly every day. Participants were categorized as having depressive symptoms (score≥10) or not (score<10) [25]. This cutoff provides reliable sensitivity and specificity for the detection of major depressive disorder [25]. The PHQ-9 has also been shown to be a reliable measure of depression severity and has been validated [25]. As a sensitivity analysis, statistical analyses were conducted using continuous depressive symptom scores.

Secondary Outcomes
Secondary outcomes were cognitive-affective (sum of responses to PHQ-9 items 1, 2, 6, 7, 8, and 9) and somatic (sum of responses to PHQ-9 items 3, 4, and 5) symptoms of depression [21,26]. These symptom subgroups were analyzed on a continuous scale.

Covariates
To account for potential confounding bias in the relationship between PA and depressive symptoms, we adjusted for age [27], sex [28], race [29], education [30], marital status, socioeconomic status (ratio of family income-to-poverty threshold) [31], BMI [32], sleep time on weekdays or workdays [33], hours of sedentary activity [33], cigarette use [34], and general self-reported health condition. Age was continuous by 1-year increases. Sex was dichotomized. Race was categorized as Mexican American, non-Hispanic White, non-Hispanic Black, other Hispanic, and other race or multiracial. Education level was categorized as less than high school, high school or equivalent, some college or Associate of Arts degree, and college graduate or above. Marital status was categorized as married or living with partner, divorced or separated or widowed, and never married. The ratio of family income-to-poverty threshold was dichotomized: ≤1.3=low income and >1.3=mid-to-high income [35]. BMI was categorized as <18, 18 to <25, 25 to <30, and ≥30 kg/m² [36]. Sleep time and sedentary activity were continuous by hourly increases. General self-reported health condition was categorized as poor, fair, and good or above. Cigarette use was dichotomized; individuals who answered “Yes” to “Smoked at least 100 cigarettes in life” and answered “Every day” or “Some days” to “Do you now smoke cigarettes?” on the Smoking-Cigarette Use Questionnaire were considered cigarette users, and those who answered “No” or “Not at all” to the respective questions were considered nonusers.

Statistical Analysis
Statistical analyses were performed using R (version 4.2.1, R Foundation for Statistical Computing) and the package survey to account for the NHANES survey design. Mobile Examination Center survey weights were used and divided by 6 to account for merging 6 survey cycles. Weighted means and SDs were estimated for continuous variables, and weighted proportions with unweighted frequencies were calculated for categorical variables. Continuous variables were compared using 2-tailed t tests, while categorical variables were compared using chi-square tests. The primary and secondary analyses for the outcome of depressive symptoms were performed using multivariable logistic regression models while adjusting for covariates. The exploratory analysis for the secondary outcomes, namely cognitive-affective and somatic symptoms scores, used multivariable linear regression models, also adjusting for covariates. The backward stepwise selection was used to develop the final multivariable models with null models including age, sex, race, and socioeconomic status. Statistical significance was established as P≤.05, and P values were adjusted using the Holm method to account for multiple comparisons. Missing data were accounted for using an available case analysis. The sample size
depended on data availability and as such, no a priori power calculations were performed.

### Results

#### Descriptive Statistics

The final study sample consisted of 29,730 participants (51.34%, n=15,133, were female) aged 20-80 (mean 47.62, SD 16.99) years (refer to Figure 1 for a detailed breakdown of participant inclusion). Depressive symptom prevalence was 2739 (8.07%) participants. The descriptive statistics for each measure can be found in Table 1.

#### Figure 1.
Flowchart of inclusion in the cross-sectional analyses; participants from a nationally representative sample of the United States obtained through the NHANES, 2007-2018. DPQ: Mental Health—Depression Screener; NHANES: National Health and Nutrition Examination Survey; PAQ: Physical Activity Questionnaire.
Table 1. Characteristics of included participants in the cross-sectional analyses from a nationally representative sample of the United States obtained through the National Health and Nutrition Examination Survey, 2007-2018.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No depressive symptoms (&lt;10; n=26,991)</th>
<th>Depressive symptoms (≥10; n=2739)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>47.68 (17.08)</td>
<td>46.99 (15.94)</td>
<td>.15</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female</td>
<td>13,385 (50.18)</td>
<td>1748 (64.54)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13,606 (49.82)</td>
<td>991 (35.46)</td>
<td></td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mexican American</td>
<td>4059 (8.52)</td>
<td>405 (8.03)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>5775 (10.85)</td>
<td>588 (13.27)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>11,210 (67.39)</td>
<td>1157 (63.31)</td>
<td></td>
</tr>
<tr>
<td>Other Hispanic</td>
<td>2724 (5.59)</td>
<td>377 (8.01)</td>
<td></td>
</tr>
<tr>
<td>Other race or multiracial</td>
<td>3223 (7.66)</td>
<td>212 (7.38)</td>
<td></td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Less than high school</td>
<td>6107 (14.41)</td>
<td>979 (25.51)</td>
<td></td>
</tr>
<tr>
<td>High school or GED</td>
<td>6173 (22.82)</td>
<td>653 (26.54)</td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>8047 (31.4)</td>
<td>819 (33.78)</td>
<td></td>
</tr>
<tr>
<td>College and above</td>
<td>6643 (31.37)</td>
<td>285 (14.17)</td>
<td></td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Married or living with other</td>
<td>16,366 (64.59)</td>
<td>1208 (46.9)</td>
<td></td>
</tr>
<tr>
<td>Divorced or separated or widowed</td>
<td>5718 (17.34)</td>
<td>948 (31.22)</td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>4892 (18.07)</td>
<td>580 (21.88)</td>
<td></td>
</tr>
<tr>
<td>PIRc, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≤1.3 (low income)</td>
<td>7335 (19.84)</td>
<td>1319 (41.89)</td>
<td></td>
</tr>
<tr>
<td>&gt;1.3 (mid-to-high income)</td>
<td>17,248 (80.16)</td>
<td>1153 (58.11)</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²), n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;18</td>
<td>392 (1.42)</td>
<td>50 (2.12)</td>
<td></td>
</tr>
<tr>
<td>18 to &lt;25</td>
<td>7259 (27.93)</td>
<td>590 (23.9)</td>
<td></td>
</tr>
<tr>
<td>25 to &lt;30</td>
<td>8975 (33.49)</td>
<td>708 (26.09)</td>
<td></td>
</tr>
<tr>
<td>≥30</td>
<td>10,112 (37.16)</td>
<td>1350 (47.89)</td>
<td></td>
</tr>
<tr>
<td>Cigarette use, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>5054 (17.85)</td>
<td>1014 (39.37)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>21,920 (82.15)</td>
<td>1725 (60.63)</td>
<td></td>
</tr>
<tr>
<td>Health condition, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Good or above</td>
<td>21,403 (85.17)</td>
<td>1111 (49.08)</td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>4989 (13.33)</td>
<td>1111 (36.19)</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>597 (1.50)</td>
<td>517 (14.73)</td>
<td></td>
</tr>
<tr>
<td>PAd (hours per week), mean (SD)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Total PA</td>
<td>10.87 (16.85)</td>
<td>8.82 (16.27)</td>
<td></td>
</tr>
<tr>
<td>Moderate-intensity PA</td>
<td>6.81 (10.98)</td>
<td>5.71 (10.83)</td>
<td>.001</td>
</tr>
<tr>
<td>Vigorous-intensity PA</td>
<td>4.06 (9.27)</td>
<td>3.10 (8.50)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Recreational PA</td>
<td>2.72 (4.46)</td>
<td>1.52 (4.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Work-related PA</td>
<td>8.15 (15.83)</td>
<td>7.30 (15.19)</td>
<td>.06</td>
</tr>
</tbody>
</table>
**Weekly Duration of Total PA and Depressive Symptoms**

After controlling for all covariates, for every 1-hour increase in the weekly duration of total PA, there was a nonsignificant decrease in the odds of depressive symptoms (Table 2) and a decrease in the somatic and cognitive-affective scores of depression (Table 3). The association with weekly duration of total PA remained nonsignificant in the sensitivity analysis (Table S2 in Multimedia Appendix 1).

### Table 2. Weighted ORs and aORs for associations between PA and depressive symptoms for the cross-sectional analyses of participants from a nationally representative sample of the United States obtained through the National Health and Nutrition Examination Survey, 2007-2018.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No depressive symptoms (&lt;10; n=26,991)</th>
<th>Depressive symptoms (≥10; n=2739)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate-intensity work-related PA</td>
<td>5.17 (10.27)</td>
<td>4.64 (9.98)</td>
<td>.07</td>
</tr>
<tr>
<td>Vigorous-intensity work-related PA</td>
<td>2.98 (8.73)</td>
<td>2.66 (8.09)</td>
<td>.17</td>
</tr>
<tr>
<td>Moderate-intensity recreational PA</td>
<td>1.63 (3.15)</td>
<td>1.08 (3.09)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vigorous-intensity recreational PA</td>
<td>1.08 (2.63)</td>
<td>0.44 (1.92)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>PHQ-9a, mean (SD)</td>
<td>2.13 (2.39)</td>
<td>14.06 (3.81)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Cognitive-affective, mean (SD)</td>
<td>0.78 (1.34)</td>
<td>7.85 (3.27)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Somatic, mean (SD)</td>
<td>1.35 (1.55)</td>
<td>6.21 (1.94)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aUnweighted frequencies are paired with weighted percentages. Mean and SDs are weighted. P values are based on weighted values.

bGED: General Educational Development.
cPIR: poverty-income ratio.
dPA: physical activity.
ePHQ-9: Patient Health Questionnaire-9.

---

**Weekly Duration of Total PA and Depressive Symptoms**

After controlling for all covariates, for every 1-hour increase in the weekly duration of total PA, there was a nonsignificant decrease in the odds of depressive symptoms (Table 2) and a decrease in the somatic and cognitive-affective scores of depression (Table 3). The association with weekly duration of total PA remained nonsignificant in the sensitivity analysis (Table S2 in Multimedia Appendix 1).

### Table 2. Weighted ORs and aORs for associations between PA and depressive symptoms for the cross-sectional analyses of participants from a nationally representative sample of the United States obtained through the National Health and Nutrition Examination Survey, 2007-2018.

<table>
<thead>
<tr>
<th>PA</th>
<th>Depressive symptoms (PHQ-9≥10)</th>
<th>P value</th>
<th>aOR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total PA</td>
<td>0.992 (0.987-0.996)</td>
<td>.002</td>
<td>0.996 (0.991-1.000)</td>
<td>.20</td>
</tr>
<tr>
<td>Moderate-intensity PA</td>
<td>0.990 (0.983-0.997)</td>
<td>.007</td>
<td>0.996 (0.989-1.003)</td>
<td>.54</td>
</tr>
<tr>
<td>Vigorous-intensity PA</td>
<td>0.986 (0.978-0.995)</td>
<td>.005</td>
<td>0.991 (0.983-0.999)</td>
<td>.14</td>
</tr>
<tr>
<td>Recreational PA</td>
<td>0.902 (0.877-0.928)</td>
<td>&lt;.001</td>
<td>0.965 (0.944-0.986)</td>
<td>.007</td>
</tr>
<tr>
<td>Work-related PA</td>
<td>0.996 (0.992-1.000)</td>
<td>.08</td>
<td>0.997 (0.993-1.002)</td>
<td>.54</td>
</tr>
</tbody>
</table>

aOR: odds ratio.
aOR: adjusted odds ratio.
PA: physical activity.

P<.05 indicates statistical significance, and P values are adjusted using the Holm method. Adjusted models adjusted for the following covariates: age, sex, race, education, marital status, socioeconomic status (ratio of family income-to-poverty threshold), BMI, sleep time on weekdays or workdays, hours of sedentary activity, cigarette use, and general self-reported health condition.

PHQ-9: Patient Health Questionnaire-9.
Table 3. Associations between PA\textsuperscript{a} and somatic and cognitive-affective symptoms of depression for the cross-sectional analyses of participants from a nationally representative sample of the United States obtained through the National Health and Nutrition Examination Survey, 2007-2018\textsuperscript{b}.

<table>
<thead>
<tr>
<th>PA</th>
<th>Somatic symptoms</th>
<th>Cognitive-affective symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\beta) (95% CI)</td>
<td>(P) value</td>
</tr>
<tr>
<td>Total PA</td>
<td>(-0.001 (-0.003 to 0.001))</td>
<td>(\geq.99)</td>
</tr>
<tr>
<td>Moderate-intensity PA</td>
<td>0.000 (-0.004 to 0.004)</td>
<td>(\geq.99)</td>
</tr>
<tr>
<td>Vigorous-intensity PA</td>
<td>(-0.003 (-0.006 to 0.001))</td>
<td>(0.41)</td>
</tr>
<tr>
<td>Recreational PA</td>
<td>(-0.016 (-0.022 to -0.009))</td>
<td>(&lt;.001)</td>
</tr>
<tr>
<td>Work-related PA</td>
<td>0.000 (-0.002 to 0.003)</td>
<td>(\geq.99)</td>
</tr>
</tbody>
</table>

\(a\): physical activity.

\(b\): \(P\leq.05\) indicates statistical significance, and \(P\) values are adjusted using the Holm method. Adjusted models adjusted for the following covariates: age, sex, race, education, marital status, socioeconomic status (ratio of family income-to-poverty threshold), BMI, sleep time on weekdays or workdays, hours of sedentary activity, cigarette use, and general self-reported health condition.

\(c\): \(\beta\): adjusted coefficient.

**Weekly Duration of PA, Subset by Intensity, and Depressive Symptoms**

After adjusting for all covariates, for every 1-hour increase in the weekly duration of moderate- and vigorous-intensity PA, there was no significant association with the odds of depressive symptoms (all \(P>.05\); Table 2) and the somatic and cognitive-affective scores of depression (all \(P>.05\); Table 3). The association with weekly duration of moderate- and vigorous-intensity PA remained nonsignificant in the sensitivity analysis with the outcome of depressive symptoms (all \(P>.05\); Table S2 in Multimedia Appendix 1).

**Weekly Duration of PA, Subset by Domain, and Depressive Symptoms**

After controlling for the covariates, for every 1-hour increase in the weekly duration of recreational PA, there was a significant decrease in the odds of depressive symptoms (aOR 0.965, 95% CI 0.944-0.986; \(P=.007\); Table 2) and the somatic and cognitive-affective scores of depression (aOR –0.015, 95% CI –0.023 to –0.007; \(P=.003\); Table 3). The association with the weekly duration of recreational PA remained statistically significant in the sensitivity analysis (\(P<.001\); Table S2 in Multimedia Appendix 1).

Covariate-adjusted models showed that there was a nonsignificant relationship between increased weekly duration of work-related PA and decreased odds of depressive symptoms (\(P=.54\); Table 2). Furthermore, for every 1-hour increase in the weekly duration of work-related PA, there was no significant association with cognitive-affective and somatic scores of depression (\(P=.36\); Table 3). The association with weekly duration of work-related PA remained nonsignificant in the sensitivity analysis (\(P>.99\); Table S2 in Multimedia Appendix 1).

**Weekly Duration of PA, Subset by Domain and Intensity, and Depressive Symptoms**

A 1-hour increase in the weekly duration of vigorous-intensity recreational PA (aOR 0.926, 95% CI 0.883-0.972; \(P=.009\)) was significantly associated with 7.4% lower odds of having depressive symptoms, but there was no significant association between depressive symptom odds and the weekly duration of moderate-intensity recreational PA (aOR 0.975, 95% CI 0.949-1.001; \(P=.19\); Table 4). The association with the weekly duration of vigorous-intensity recreational PA (\(P=.001\)) and moderate-intensity recreational PA (\(P=.007\)) was statistically significant in the sensitivity analysis with the outcome of depressive symptoms (Table S3 in Multimedia Appendix 1).

For every 1-hour increase in the weekly duration of vigorous-intensity recreational PA, the somatic (adjusted coefficient [\(\beta\]=–0.021, 95% CI –0.032 to –0.010; \(P=.001\)) and cognitive-affective (\(\beta\)=–0.022, 95% CI –0.035 to –0.009; \(P=.005\)) scores of depression decreased significantly, but moderate-intensity recreational PA was only associated with significant decreases in somatic scores of depression (\(\beta\)=–0.016, 95% CI –0.025 to –0.007; \(P=.002\)). No significant associations were found for the weekly duration of vigorous-intensity work-related PA or moderate-intensity work-related PA and depressive symptoms (all \(P>.05\); Table 5). The association with weekly duration of vigorous-intensity work-related PA or moderate-intensity work-related PA remained nonsignificant in the sensitivity analysis (all \(P>.05\); Table S3 in Multimedia Appendix 1).
This study investigated the association of the weekly duration of PA and depressive symptoms. Depressive symptom odds decreased significantly with an increase in the weekly duration of recreational PA. When intensity was examined alone, neither vigorous- nor moderate-intensity PA was significantly associated with depressive symptoms. When domain and intensity were considered together, only an increase in vigorous-intensity recreational PA duration was significantly associated with a decrease in the odds of depressive symptoms. Notably, based on the demographic data collected in this study, participants with depressive symptoms were more likely to be female, have a higher BMI, have lower income, and participated in lower weekly durations of PA regardless of being subset by intensity and domain. This coincides with prior studies that have shown gender [37], BMI [32], income status [38], and PA to be associated with the likelihood of depression. As a result, to address potential confounding bias, we incorporated these variables as covariates when examining the association between the weekly duration of PA and depressive symptoms.

Our findings coincide with results from previous studies that found a significant association between decreased recreational PA and greater depressive symptoms. In addition to the neurobiological changes, such as an increased release of endorphins and neurotransmitters like dopamine and norepinephrine [39-41], the protective effect of recreational PA involves social aspects and enjoyability of recreational activities [14]. Recreational activities can be social or self-focused, both providing time structure [42-44]. Social recreational PA offers an added benefit of pleasurable social interactions with others, which can provide a distraction from negative life events and increase perceptions of social support [43,44]. Alternatively, self-focused recreational PA can be conducted individually or within a group setting and buffers against negative events through personal transformation [43]. Thus, it may be expected that recreational PA has a more comprehensive effect on

Table 4. Weighted ORs and aORs for associations between PA subcategories and depressive symptoms for the cross-sectional analyses of participants from a nationally representative sample of the United States obtained through the National Health and Nutrition Examination Survey, 2007-2018.

<table>
<thead>
<tr>
<th>PA</th>
<th>Depressive symptoms (PHQ-9≥10)</th>
<th></th>
<th>aOR (95% CI)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>P value</td>
<td></td>
<td>P value</td>
</tr>
<tr>
<td>Moderate-intensity work-related PA</td>
<td>0.995 (0.988-1.001)</td>
<td>.19</td>
<td>0.998 (0.990-1.005)</td>
<td>.53</td>
</tr>
<tr>
<td>Vigorous-intensity work-related PA</td>
<td>0.995 (0.989-1.002)</td>
<td>.19</td>
<td>0.995 (0.987-1.002)</td>
<td>.35</td>
</tr>
<tr>
<td>Moderate-intensity recreational PA</td>
<td>0.916 (0.882-0.952)</td>
<td>&lt;.001</td>
<td>0.975 (0.949-1.001)</td>
<td>.19</td>
</tr>
<tr>
<td>Vigorous-intensity recreational PA</td>
<td>0.829 (0.783-0.879)</td>
<td>&lt;.001</td>
<td>0.926 (0.883-0.972)</td>
<td>.009</td>
</tr>
</tbody>
</table>

aOR: odds ratio.
b aOR: adjusted odds ratio.
c PA: physical activity.
d P<.05 indicates statistical significance, and P values are adjusted using the Holm method. Adjusted models adjusted for the following covariates: age, sex, race, education, marital status, socioeconomic status (ratio of family income-to-poverty threshold), BMI, sleep time on weekdays or workdays, hours of sedentary activity, cigarette use, and general self-reported health condition.
e PHQ-9: Patient Health Questionnaire-9.

Table 5. Associations between PA subcategories and somatic and cognitive-affective symptoms of depression for the cross-sectional analyses of participants from a nationally representative sample of the United States obtained through the National Health and Nutrition Examination Survey, 2007-2018.

<table>
<thead>
<tr>
<th>PA</th>
<th>Somatic symptoms</th>
<th></th>
<th>Cognitive-affective symptoms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aβ (95% CI)</td>
<td>P value</td>
<td>aβ (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>Moderate-intensity work-related PA</td>
<td>0.002 (–0.002 to 0.006)</td>
<td>.79</td>
<td>–0.001 (–0.004 to 0.002)</td>
<td>.55</td>
</tr>
<tr>
<td>Vigorous-intensity work-related PA</td>
<td>–0.001 (–0.005 to 0.003)</td>
<td>.79</td>
<td>–0.003 (–0.007 to 0.001)</td>
<td>.34</td>
</tr>
<tr>
<td>Moderate-intensity recreational PA</td>
<td>–0.016 (–0.025 to –0.007)</td>
<td>.002</td>
<td>–0.014 (–0.026 to –0.003)</td>
<td>.05</td>
</tr>
<tr>
<td>Vigorous-intensity recreational PA</td>
<td>–0.021 (–0.032 to –0.010)</td>
<td>.001</td>
<td>–0.022 (–0.035 to –0.009)</td>
<td>.005</td>
</tr>
</tbody>
</table>

a PA: physical activity.
b P<.05 indicates statistical significance, and P values are adjusted using the Holm method. Adjusted models adjusted for the following covariates: age, sex, race, education, marital status, socioeconomic status (ratio of family income-to-poverty threshold), BMI, sleep time on weekdays or workdays, hours of sedentary activity, cigarette use, and general self-reported health condition.
c aβ: adjusted coefficient.
depression with reduced cognitive-affective and somatic depressive symptoms. Our results demonstrated that recreational PA has similar associations with both symptom subgroups, suggesting that its impact is not limited to changes in one’s psychosocial experience. The influence on somatic symptoms may be due to recreational PA frequently occurring outside, being more enjoyable, and producing euphoric feelings that are captured as somatic [45].

In examining the statistically significant associations between the different aspects of PA and depressive symptoms, it is important to note the statistical issue that the increased power in a large sample study leads to narrower CIs for the effect measures and smaller P values. However, by adjusting the statistical significance using the Holm method, the chances of false positives were minimized. Acknowledging that this can still cause one to claim impractical significance for very small effects, we prioritized clinical significance over statistical significance.

While work-related PA has been associated with higher levels of depression [17], our results did not reach statistical significance. This could be attributed to sample differences, as previous research examined adults aged 50 and older [17], whereas our sample size consisted of participants aged 20-80 years. Since most work-related PAs entail heavy lifting and strenuous overuse of the neck and back without adequate rest and recovery [46], it may act as a burden for all adults, including older adults, and contribute to higher depression levels. As this study included both younger and older individuals aged 20-80 years, the impact of work-related PA on depressive symptoms may be influenced by additional factors such as job satisfaction and job type [47,48]. Additionally, in this study, work-related PA encompassed paid and unpaid work, household chores, and yard work. Given that research has found a greater frequency of domestic activities, such as housework and gardening, to be associated with lower odds of psychological distress [12], our findings should be interpreted with caution, as the 2 types of work cannot be distinguished based on the NHANES questions. Furthermore, it is important to explore an individual’s interest and competence in executing workplace PAs. Prior research suggests that within the work domain, PA may be associated with limited personal choice and is typically driven by the demands of work output and the flow of work, which can be strenuous [49]. Another study highlights the significance of occupation type, shedding light on how work-related PA can vary across different job categories [48]. For example, women in trade, labor, or transport-related jobs who engage in work-related walking experienced lower levels of psychological distress, but this association was not present when considering moderate or vigorous PAs [48]. This suggests that the nature of the job and the type of PA involved are linked to mental health outcomes. Further investigation into this will help parse out the mechanism by which work-related PA is associated with depressive symptoms.

Regular PA participation can provide a sense of mastery, improved body image, sense of achievement, and feelings of control that may distract one from negative thoughts [16]. However, this study found no significant association between the weekly duration of total PA and depressive symptom odds, which challenges previous literature that found a greater frequency of PA has a protective effect on depressive symptoms [50,51]. This contradiction may be attributable to the absence of low-intensity PA duration in the total PA calculations due to its unavailability in the NHANES data. As low-intensity PAs involve reduced energy expenditure than moderate or vigorous-intensity PA [7], they may be easier to engage in more frequently and thus show important associations with depressive symptoms. In fact, previous studies have found that light and moderate intensities of PA were both associated with lower likelihoods of depression [8,52].

Contrary to previous research [53,54], we found that hourly increases in the weekly duration of moderate- and vigorous-intensity PA did not significantly decrease depressive symptom odds. Previous research shows intensity-dependent physiological responses caused by increased frequencies of higher-intensity PA, all of which correlate with reduced depressive symptoms [55]. These physiological responses include the adaptation of the hypothalamic-pituitary-adrenal axis activity leading to increased stress resiliency [56], increased neurogenesis from elevated brain-derived neurotrophic factor expression [57], and reduced inflammation creating an anti-inflammatory environment [40,55]. The discrepancy between our study and prior literature may be attributed to the fact that our PA variables are defined using weekly duration rather than frequency. Therefore, participants with the same weekly duration of moderate-intensity or vigorous-intensity PA may be engaging in a different PA frequency (eg, 50 minutes per day for 3 days vs 30 minutes per day for 5 days). Prior research has shown that <10 minutes of exercise contributes to minimal mood improvements, whereas a longer duration (>30 minutes) can contribute to fatigue and withdrawal-related responses [58].

**Limitations**

This study is not without limitations. The exposure and outcome variables were measured using self-report questionnaires, which are subject to participant biases, including recall, social desirability, and nonresponse bias. This study also used secondary data and was subject to limited predefined variables. As such, although data regarding PA domains were available, we did not have information on respondents’ activities within recreational or work-related PA. Recreational PA encompasses sports, fitness, and brisk walking, whereas work-related PA encompasses paid, unpaid work, and domestic activities. Additionally, our analysis did not consider light-intensity PA since it was not available in the NHANES questionnaire. Finally, the cross-sectional design of the study limits the conclusions to correlational rather than causal. There is a bidirectional association between PA and depression, wherein it is possible that there is a reverse causal relationship called the inhibition hypothesis [59,60], which suggests that depression symptoms such as anhedonia, low mood, and lack of energy may inhibit individuals from engaging in PA due to a lack of interest or motivation [59].

**Conclusions and Future Directions**

This study demonstrates that PA domain, specifically recreational PA, might play an integral role in the protective
effect of PA on depressive symptoms. Future studies should further investigate the impact of different types of recreational PA (social and self-focused) and their impact on depressive symptoms. While we found that an increased weekly duration of work-related PA was not significantly associated with lower depressive symptoms, future studies should explore the different types of occupational and household-based PA to further elucidate which activities have a detrimental effect and which have a beneficial effect. Furthermore, occupational PA can be further subdivided into paid and unpaid work to examine their effects on depressive symptoms, as these are likely to be driven by different motivations.

Acknowledgments
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability
The data sets analyzed during this study are available via the NHANES website [61].

Authors’ Contributions
JKB and VB conceptualized the study. The investigation was led by JKB, who was also responsible for writing the original draft and visualization. Study conceptualization and manuscript writing were supervised by VB, VKT, and SD. MW and ZC were responsible for the methodology, data curation, and formal analysis. WL and HJ supervised formal analysis. All authors provided critical revision of the manuscript for important intellectual content and contributed to editing. All authors had full access to the data and read and approved the final manuscript.

Conflicts of Interest
JKB, SD, MW, VKT, SFD, VZC, ZC, HJ, and WL do not have any disclosures. VB is supported by an Academic Scholar Award from the University of Toronto Department of Psychiatry and has received research support from the Canadian Institutes of Health Research, Brain & Behaviour Foundation, Ontario Ministry of Health Innovation Funds, Royal College of Physicians and Surgeons of Canada, Department of National Defence (Government of Canada), New Frontiers in Research Fund, Associated Medical Services Inc Healthcare, American Foundation for Suicide Prevention, Roche Canada, Novartis, and Eisai.

Multimedia Appendix 1
Detailed methodology and statistical analysis results.

References
44. Goodman WK, Geiger AM, Wolf JM. Leisure activities are linked to mental health benefits by providing time structure: comparing employed, unemployed and homemakers. J Epidemiol Community Health 2017;71(1):4-11 [FREE Full text] [doi: 10.1136/jech-2016-207260] [Medline: 27298424]
Abbreviations

- $\beta$: adjusted coefficient
- aOR: adjusted odds ratio
- CDC: Centers for Disease Control and Prevention
- NCHS: National Center for Health Statistics
- NHANES: National Health and Nutrition Examination Survey
- OR: odds ratio
- PA: physical activity
- PAQ: Physical Activity Questionnaire
- PHQ-9: Patient Health Questionnaire-9

Edited by T de Azvedo Cardoso; submitted 26.04.23; peer-reviewed by B Ainsworth, S Lopes; comments to author 24.10.23; revised version received 14.11.23; accepted 20.05.24; published 05.07.24.

Please cite as:


The Association Between Depressive Symptoms and the Weekly Duration of Physical Activity Subset by Intensity and Domain: Population-Based, Cross-Sectional Analysis of the National Health and Nutrition Examination Survey From 2007 to 2018


URL: https://www.i-jmr.org/2024/1/e48396
doi:10.2196/48396
PMID:
The Effect of Body Temperature Changes on the Course of Treatment in Patients With Pneumonia and Sepsis: Results of an Observational Study

Domen Guzelj1*; Anže Grubelnik1*; Nina Greif1*; Petra Povalej Bržan2*, PhD; Jure Fluher3*, MD; Žiga Kalamar3*, MD; Andrej Markota3*, MD, PhD

1Faculty of Medicine, University of Maribor, Maribor, Slovenia
2Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia
3Medical Intensive Care Unit, University Medical Centre Maribor, Maribor, Slovenia

*all authors contributed equally

Corresponding Author:
Andrej Markota, MD, PhD
Medical Intensive Care Unit
University Medical Centre Maribor
Ljubljanska ulica 5
Maribor, 2000
Slovenia
Phone: 386 23212008
Email: andrej.markota@ukc-mb.si

Abstract

Background: Traditionally, patients who are critically ill with infection and fever have been treated with antipyretics or even physically cooled. Presumed benefits of the reduction of body temperature are mostly based on decreased metabolic demands. However, it has been shown that decreasing body temperature in patients who are critically ill is not associated with improvement in treatment outcomes. Additionally, there is some data to support the use of temperature modulation (therapeutic hyperthermia) as an adjuvant treatment strategy in patients with infection.

Objective: This study aims to determine the effect of body temperature on the course of intensive care unit (ICU) treatment of patients who are mechanically ventilated with pneumonia, sepsis, and positive tracheal aspirates on admission.

Methods: We performed a single-center retrospective study. Core body temperature was measured in all patients. We analyzed associations between average temperatures in the first 48 hours after admission to ICU and ICU treatment parameters. Additionally, patients were divided into three groups: patients with negative tracheal aspirates 1 week after ICU admission (P-N group), patients with a different pathogen in tracheal aspirates 1 week after ICU admission (P-HAP group), and patients with a persisting pathogen in tracheal aspirates 1 week after ICU admission (P-P group). Differences in body temperature and interventions aimed at temperature modulation were determined.

Results: We observed a significantly higher average temperature in the first 48 hours after admission to ICU in patients who survived to hospital discharge compared to nonsurvivors (mean 37.2 °C, SD 1 °C vs mean 36.9 °C, SD 1.6 °C; P=.04). We observed no associations between average temperatures in the first 48 hours after ICU admission and days of mechanical ventilation in the first 7 days of treatment (ρ=-0.090; P=.30), the average maximum daily requirement for noradrenaline in the first 7 days of treatment (ρ=-0.029; P=.80), average maximum FiO2 in the first 7 days of ICU treatment (ρ=0.040; P=.70), and requirement for renal replacement therapy in the first 7 days of ICU treatment (mean 37.3 °C, SD 1.4 °C vs mean 37.0 °C, SD 1.3 °C; P=.23). In an additional analysis, we observed a significantly greater use of paracetamol in the P-N group (mean 1.0, SD 0.7 g vs mean 0.4, SD 0.8 g; P=.009), a trend toward greater use of active cooling in the first 24 hours after ICU admission in the P-N group (n=11, 44% vs n=14, 33.3% vs n=16, 32%; P=.57), and no other significant differences in parameters of ICU treatment between patient groups.

Conclusions: We observed better survival in patients who developed higher body temperatures in the first 48 hours after admission to the ICU; however, we observed no changes in other treatment parameters. Similarly, we observed greater use of...
paracetamol in patients with negative tracheal aspirates 1 week after ICU admission. Our results support the strategy of temperature tolerance in patients who are intubated with pneumonia and sepsis.


KEYWORDS
fever; targeted temperature management; pneumonia; sepsis; intensive care unit

Introduction
Increased body temperature has been recognized as a sign of illness for more than 2000 years and antipyretics have been used for at least 100 years with the aim of lowering body temperature in patients who are febrile [1]. However, a body of evidence exists ranging from studies in the preantibiotic era to a recently published pilot trial, where increased body temperature is investigated as a treatment option in patients with infection [2-4]. A fever-range increase of body temperature is a highly preserved response that is probably beneficial in patients who develop fever as a part of a normal immune response to infection [5]. Recently, we have observed that patients with lower body temperatures are at a higher risk of acquiring the presence of multidrug-resistant pathogens [6]. The aim of this study was to investigate the association between body temperature and the course of intensive care unit (ICU) treatment in patients with pneumonia and sepsis and to evaluate the effect of temperature on the persistence of positive control tracheal aspirates.

Methods

Study Design and Settings
We performed retrospective observational data collection from January 1, 2018, to December 31, 2021. The study was performed in a medical ICU in a tertiary center and supported by an institutional research grant (grant IRP-2022/01-01).

Ethical Considerations
The study was approved by University Medical Centre Maribor Ethical Committee, and informed consent was waived because of the observational and retrospective nature of the study (No. UKC-MB-KME-35/22). Patients’ data have been anonymized and deidentified. No compensation was provided to the participants.

Study Population
We included adult (aged >18 years) patients with pneumonia, septic shock, positive tracheal aspirates within 24 hours after admission (“admission” tracheal aspirates), and tracheal aspirates withdrawn 8-14 days after admission to ICU (“week 2” tracheal aspirates). We excluded patients who were treated with targeted temperature management for accidental hypo- or hyperthermia, patients after cardiac arrest, and patients who are neurocritical (eg, patients with meningitis, encephalitis, ischemic or hemorrhagic stroke, or subarachnoid hemorrhage). Temperature management in our patient cohort was as per the treating physician.

Measurments
We collected basic demographic data and data related to ICU treatment, namely outcome of ICU treatment, ICU length of stay, core body temperature during ICU stay, use of renal replacement therapy, use of acetaminophen, maximum concentration of noradrenaline, maximum fraction of inspired oxygen, maximum level of positive end-expiratory pressure and maximum minute ventilation, and microbiological results of tracheal aspirates. Source data was paper based for temperature and therapeutic charts and electronic for other data. Core body temperature was used for study data. As per department policy, temperature was measured via thermal probe urinary catheters (Rüscher Sensor Urinary Catheters, Teleflex Medical, Athlone, Ireland), and temperature measurements were continuously displayed on ICU monitors (Philips IntelliVue MX800 Patient Monitoring System, Koninklijke Philips N.V., Amsterdam, Netherlands). If insertion of a urinary catheter is not possible, then an esophageal temperature probe is inserted, but temperature measurement in all of the included patients was performed via a urinary catheter. Temperature results from 2-hourly notations were used for statistical analysis. We compared the association between body temperature and course of treatment parameters between three groups of patients: patients with sterile week 2 tracheal aspirates (P-P group), patients with a different pathogen in week 2 tracheal aspirates (P-HAP group), and patients with a persistent pathogen presence in week 2 tracheal aspirates (P-P group).

Data Analysis
Statistical analyses were carried out using R (version 4.1.1: R Foundation for Statistical Computing). Nominal variables are presented with frequencies (percentages) and numerical variables with means (SDs) or medians (IQRs) when the normality assumption is violated. For the comparison of nominal dichotomous variables, the Fisher exact test was used. Continuous variables were first assessed for normality using the D’Agostino omnibus test. A comparison of continuous variables across groups was carried out using the Kruskal-Wallis test. Dunn post hoc test with Bonferroni correction was used to adjust for multiple comparisons. The association between average temperature in the first 48 hours and the duration of mechanical ventilation, the average maximum daily requirement for noradrenaline and the average maximum FiO₂ in the first 7 days was evaluated using Spearman correlation. Generalized linear models were used to additionally estimate the ICU and hospital survival in association with the abovementioned risk factors. A statistically significant observation was considered at P<.05.
**Results**

**Baseline Characteristics**

In all, 117 patients were included in the study analysis; 84 (71.8%) were male, and the mean age was 63.7 (SD 13.5) years. All patients were invasively mechanically ventilated in the ICU on day 1. Mean APACHE II and SOFA scores on admission were 21.7 (SD 6.6) and 10.1 (SD 2.6), respectively. A total of 77 (65.8%) were discharged alive from the ICU, and 46 (39.3%) patients were discharged alive from the hospital. P-N, P-HAP, and P-P groups consisted of 25, 42, and 50 patients, respectively. We observed no significant differences between the P-N, P-HAP, and P-P groups in all treatment parameters apart from the use of paracetamol on day 1, which was significantly greater in the P-N group (mean 1.0, SD 1.1 g vs mean 0.4, SD 0.7 g vs mean 0.4, SD 0.8 g; $P=0.009$). General demographic data and parameters describing the course of the treatment in the ICU are described in Table 1. The study patient population and inclusion flowchart are presented in Figure 1.

Table 1. Demographic data and parameters describing the course of treatment in the intensive care unit (ICU).

<table>
<thead>
<tr>
<th>Label and variable</th>
<th>Group</th>
<th>Total</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients, n (%)</td>
<td>P-N$^a$</td>
<td>P-P$^b$</td>
<td>P-HAP$^c$</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5 (15.2)</td>
<td>18 (54.5)</td>
<td>10 (30.3)</td>
</tr>
<tr>
<td>Male</td>
<td>20 (23.8)</td>
<td>32 (38.1)</td>
<td>32 (38.1)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>62.0 (52.0-74.0)</td>
<td>64.0 (57.2-71.8)</td>
<td>67.0 (62.0-76.0)</td>
</tr>
<tr>
<td>Min-max</td>
<td>23.0-84.0</td>
<td>26.0-84.0</td>
<td>39.0-86.0</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>28.7 (7.0)</td>
<td>27.6 (4.3)</td>
<td>29.7 (6.4)</td>
</tr>
<tr>
<td>Min-max</td>
<td>17.2-47.7</td>
<td>18.8-39.2</td>
<td>16.3-44.8</td>
</tr>
<tr>
<td>Length of hospitalization (days)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>22.0 (17.0-29.0)</td>
<td>19.5 (15.0-26.8)</td>
<td>17.5 (13.0-26.8)</td>
</tr>
<tr>
<td>Min-max</td>
<td>8.0-53.0</td>
<td>10.0-43.0</td>
<td>8.0-69.0</td>
</tr>
<tr>
<td>ICU survival, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Died</td>
<td>6 (15.0)</td>
<td>16 (40.0)</td>
<td>18 (45.0)</td>
</tr>
<tr>
<td>Survived</td>
<td>19 (26.1)</td>
<td>34 (44.2)</td>
<td>24 (31.2)</td>
</tr>
<tr>
<td>Hospital survival, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Died</td>
<td>13 (18.3)</td>
<td>31 (43.7)</td>
<td>27 (38.0)</td>
</tr>
<tr>
<td>Survived</td>
<td>12 (26.1)</td>
<td>19 (41.3)</td>
<td>15 (32.6)</td>
</tr>
<tr>
<td>SOFA$^d$ at the time of admission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>9.5 (7.0-12.0)</td>
<td>10.0 (9.0-12.0)</td>
<td>10.0 (9.0-12.0)</td>
</tr>
<tr>
<td>Min-max</td>
<td>5.0-16.0</td>
<td>5.0-15.0</td>
<td>4.0-15.0</td>
</tr>
<tr>
<td>APACHE$^e$ II at the time of admission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>20.4 (6.3)</td>
<td>22.0 (6.4)</td>
<td>22.1 (7.2)</td>
</tr>
<tr>
<td>Min-max</td>
<td>6.0-32.0</td>
<td>7.0-36.0</td>
<td>5.0-40.0</td>
</tr>
<tr>
<td>Change of antibiotic in 7 days after admission (no), n (%)</td>
<td>7 (18.4)</td>
<td>17 (44.7)</td>
<td>14 (36.8)</td>
</tr>
</tbody>
</table>

$^a$P-N: patients with sterile week 2 tracheal aspirates.

$^b$P-HAP: patients with a different pathogen in week 2 tracheal aspirates.

$^c$P-P group: patients with a persistent pathogen presence in week 2 tracheal aspirates.

$^d$SOFA: Sequential Organ Failure Assessment.

$^e$APACHE: Acute Physiology and Chronic Health Evaluation.
Main Results

Body temperature changes in the first 28 days of treatment are presented in Figure 2. We observed a significantly higher average temperature in the first 48 hours after admission to ICU in patients who survived to hospital discharge compared to nonsurvivors (mean 37.2 °C, SD 1 °C vs mean 36.9 °C, SD 1.6 °C; \( P = .04 \)) and nonsignificant differences in average temperatures in the first 48 hours after admission to ICU between survivors and nonsurvivors to ICU discharge (mean 37.0 °C, SD 1.3 °C vs mean 37.0 °C, SD 1.2 °C; \( P = .60 \)). We observed no associations between average temperatures in the first 48 hours after ICU admission and days of mechanical ventilation in the first 7 days of treatment (\( \rho = -0.090; P = .30 \)), the average maximum daily requirement for noradrenaline in the first 7 days of treatment (\( \rho = -0.029; P = .80 \)), average maximum FiO\(_2\) in the first 7 days of ICU treatment (\( \rho = 0.040; P = .70 \)), and requirement for renal replacement therapy in the first 7 days of ICU treatment (mean 37.3 °C, SD 1.4 °C vs mean 37.0 °C, SD 1.3 °C; \( P = .23 \)).
Figure 2. Average body temperature by groups in the first 28 days of treatment in the intensive care unit. P-HAP: patients with a different pathogen in week 2 tracheal aspirates; P-N: patients with sterile week 2 tracheal aspirates; P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates.

P-N, P-HAP, and P-P groups consisted of 25, 42, and 50 patients each, respectively. No significant differences in average body temperature in the first 48 hours were observed between the P-P, P-HAP, and P-N groups. Temperature variations increased after day 16 (Figure 2), however, and approximately two-thirds of patients were discharged from the ICU or died before day 16. Both hospital and ICU survival were higher in the P-N group compared to the P-HAP and P-P groups; however, the differences were not statistically significant (Figure 3). We observed statistically significant greater use of paracetamol in the first 24 hours after ICU admission in the P-N group compared to the P-HAP and P-P groups (mean 1.0, SD 1.1 g vs mean 0.4, SD 0.7 g vs mean 0.4, SD 0.8 g; \( P = .009 \)). Active cooling was used more frequently on day 1 in the P-N group, but no significant differences were observed (Figure 4).

Figure 3. Intensive care unit (ICU) and hospital survival by groups. P-HAP: patients with a different pathogen in week 2 tracheal aspirates; P-N: patients with sterile week 2 tracheal aspirates; P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates.
Figure 4. Active cooling on day one of hospitalization in the intensive care unit divided by groups. P-HAP: patients with a different pathogen in week 2 tracheal aspirates; P-N: patients with sterile week 2 tracheal aspirates; P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates.

Microbiological Data
In all, 148 different causal pathogens were isolated. In admission tracheal aspirates, the majority (n=124, 66%) of pathogens were gram-negative bacteria, followed by gram-positive bacteria (n=56, 29.8%) and fungal pathogens (4.3%). In week 2 tracheal aspirates, the majority (n=128, 86.5%) of pathogens were gram-negative bacteria, followed by gram-positive bacteria (n=14, 9.5%) and fungal pathogens (n=6, 4.1%). We observed no differences between the P-N, P-HAP, and P-P groups in causative pathogens in admission tracheal aspirates (Table 2). There were no differences between the P-N, P-HAP, and P-P groups in the rate of change of antimicrobial therapy within the first week after ICU admission, and we observed no changes in the duration of initial (combined empiric and antibiogram-guided antimicrobial therapy) and cumulative ICU-stay antibiotic therapy (mean 7.6, SD 2.7 days vs mean 8.4, SD 3.1 days vs mean 8.2, SD 2.9 days; P=.55; mean 15.2, SD 5.6 days vs mean 14.8, SD 6.8 days vs mean 15.9, SD 6.2 days; P=.70; respectively).
Table 2. Isolated pathogens from the first and second samples divided by groups.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Group, n (%)</th>
<th>P-P(^a)</th>
<th>P-HAP(^b)</th>
<th>P-N(^c)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gram-positive bacteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P-P(^a)</td>
<td>P-HAP(^b)</td>
<td>P-N(^c)</td>
<td>Total, n (%)</td>
</tr>
<tr>
<td><strong>Sample 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Staphylococcus</em></td>
<td>8 (4.3)</td>
<td>15 (8.0)</td>
<td>6 (3.2)</td>
<td></td>
<td>29 (15.4)</td>
</tr>
<tr>
<td><em>Streptococcus</em></td>
<td>5 (2.7)</td>
<td>11 (5.9)</td>
<td>5 (2.7)</td>
<td></td>
<td>21 (11.2)</td>
</tr>
<tr>
<td><em>Corynebacterium</em></td>
<td>3 (1.6)</td>
<td>2 (1.1)</td>
<td>1 (0.5)</td>
<td></td>
<td>6 (3.2)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>16 (8.5)</td>
<td>28 (14.9)</td>
<td>12 (6.4)</td>
<td></td>
<td>56 (29.8)</td>
</tr>
<tr>
<td><strong>Sample 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Staphylococcus</em></td>
<td>5 (3.4)</td>
<td>2 (1.4)</td>
<td>N/A(^d)</td>
<td></td>
<td>7 (4.7)</td>
</tr>
<tr>
<td><em>Corynebacterium</em></td>
<td>1 (0.7)</td>
<td>3 (2.0)</td>
<td>N/A</td>
<td></td>
<td>4 (2.7)</td>
</tr>
<tr>
<td><em>Streptococcus</em></td>
<td>1 (0.7)</td>
<td>1 (0.7)</td>
<td>N/A</td>
<td></td>
<td>2 (1.4)</td>
</tr>
<tr>
<td><em>Enterococcus</em></td>
<td>0 (0)</td>
<td>1 (0.7)</td>
<td>N/A</td>
<td></td>
<td>1 (0.7)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>7 (4.7)</td>
<td>7 (4.7)</td>
<td>N/A</td>
<td></td>
<td>14 (9.5)</td>
</tr>
<tr>
<td><strong>Gram-negative bacteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P-P(^a)</td>
<td>P-HAP(^b)</td>
<td>P-N(^c)</td>
<td>Total, n (%)</td>
</tr>
<tr>
<td><strong>Sample 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Klebsiella</em></td>
<td>17 (9.0)</td>
<td>6 (3.2)</td>
<td>8 (4.3)</td>
<td></td>
<td>31 (16.5)</td>
</tr>
<tr>
<td><em>Haemophilus</em></td>
<td>4 (2.1)</td>
<td>7 (3.7)</td>
<td>6 (3.2)</td>
<td></td>
<td>17 (9.0)</td>
</tr>
<tr>
<td><em>Escherichia</em></td>
<td>9 (4.8)</td>
<td>4 (2.1)</td>
<td>1 (0.5)</td>
<td></td>
<td>14 (7.4)</td>
</tr>
<tr>
<td><em>Enterobacter</em></td>
<td>9 (4.8)</td>
<td>3 (1.6)</td>
<td>2 (1.1)</td>
<td></td>
<td>14 (7.4)</td>
</tr>
<tr>
<td><em>Pseudomonas</em></td>
<td>10 (5.3)</td>
<td>0 (0)</td>
<td>1 (0.5)</td>
<td></td>
<td>11 (5.9)</td>
</tr>
<tr>
<td><em>Other</em></td>
<td>23 (12.2)</td>
<td>7 (3.7)</td>
<td>7 (3.7)</td>
<td></td>
<td>37 (19.7)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>72 (38.3)</td>
<td>27 (14.4)</td>
<td>25 (13.3)</td>
<td></td>
<td>124 (66)</td>
</tr>
<tr>
<td><strong>Sample 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Klebsiella</em></td>
<td>16 (10.8)</td>
<td>12 (8.1)</td>
<td>N/A</td>
<td></td>
<td>28 (18.9)</td>
</tr>
<tr>
<td><em>Pseudomonas</em></td>
<td>16 (10.8)</td>
<td>8 (5.4)</td>
<td>N/A</td>
<td></td>
<td>24 (16.2)</td>
</tr>
<tr>
<td><em>Escherichia</em></td>
<td>11 (7.4)</td>
<td>4 (2.7)</td>
<td>N/A</td>
<td></td>
<td>15 (10.1)</td>
</tr>
<tr>
<td><em>Enterobacter</em></td>
<td>8 (5.4)</td>
<td>6 (4.1)</td>
<td>N/A</td>
<td></td>
<td>14 (9.5)</td>
</tr>
<tr>
<td><em>Acinetobacter</em></td>
<td>10 (6.8)</td>
<td>3 (2.0)</td>
<td>N/A</td>
<td></td>
<td>13 (8.8)</td>
</tr>
<tr>
<td><em>Other</em></td>
<td>18 (12.2)</td>
<td>16 (10.8)</td>
<td>N/A</td>
<td></td>
<td>34 (23.0)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>79 (53.4)</td>
<td>49 (33.1)</td>
<td>N/A</td>
<td></td>
<td>128 (86.5)</td>
</tr>
<tr>
<td><strong>Fungi</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P-P(^a)</td>
<td>P-HAP(^b)</td>
<td>P-N(^c)</td>
<td>Total, n (%)</td>
</tr>
<tr>
<td><strong>Sample 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Candida</em></td>
<td>1 (0.5)</td>
<td>6 (3.2)</td>
<td>1 (0.5)</td>
<td></td>
<td>8 (4.3)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>89 (47.3)</td>
<td>61 (32.4)</td>
<td>38 (20.2)</td>
<td></td>
<td>188 (100)</td>
</tr>
<tr>
<td><strong>Sample 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Candida</em></td>
<td>5 (3.4)</td>
<td>1 (0.7)</td>
<td>N/A</td>
<td></td>
<td>6 (4.1)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>91 (61.5)</td>
<td>57 (38.5)</td>
<td>N/A</td>
<td></td>
<td>148 (100)</td>
</tr>
</tbody>
</table>

\(^a\)P-P group: patients with a persistent pathogen presence in week 2 tracheal aspirates.

\(^b\)P-HAP: patients with a different pathogen in week 2 tracheal aspirates.

\(^c\)P-N: patients with sterile week 2 tracheal aspirates.

\(^d\)N/A: not applicable.
Discussion

Principal Findings

We observed higher initial body temperatures in patients who survived to hospital discharge and a trend toward higher initial temperatures in patients who survived to ICU discharge. We did not observe any differences in body temperature between the P-N, P-HAP, and P-P groups of patients, but we did observe a significantly greater use of paracetamol in the P-N group of patients and a trend toward greater use of active cooling in the P-N group of patients. Additionally, we observed a trend toward better survival to ICU and hospital discharge in the P-N group of patients.

Traditionally, antipyretic therapy (mostly paracetamol), and in some cases active cooling, have been used to lower body temperature in patients who are febrile with infection. The presumed patient benefit comes from decreased metabolic demands associated with lower body temperature [7,8].

The survival of our patients is comparable to studies performed by Gursel and Demirtas [9], Depuydt et al [10], and Qiao et al [11] who reported survival to ICU discharge of patients who were critically ill with pneumonia in the range between 46% and 75%, compared to 65.8% survival to ICU discharge in our patients. They included patients with similar severity of illness, with SOFA scores in the range between 4 and 6 points and APACHE II scores around 20 points, compared to the SOFA score 10.1 (SD 2.6) and APACHE II score 21.7 (SD 6.6) in our study.

Our findings are in line with a number of other studies where better survival was observed in patients with infection and physiological-grade fever (ie, body temperature between around 37 °C and 39 °C). Both Kushimoto et al [12] in a prospective observational study and Shimazui et al [13] in a retrospective observational study observed higher mortality in patients with sepsis and lower body temperature on admission. Lee et al [14] also observed lower mortality in patients with physiological-grade fever in their prospective study. Rambus et al [15] performed a meta-analysis of 42 studies reporting body temperature and mortality in patients with sepsis. They discovered a correlation between higher body temperature and lower mortality; higher body temperature predicted better survival, and hypothermia predicted lower survival. Similar to our findings the number of patients with body temperature over 39.5 °C was very low. Thomas-Rüddel et al [16] performed a secondary analysis of a large data set of patients with sepsis and discovered that initial body temperatures were distributed in two peaks: a smaller peak at around 35.5 °C and a larger peak approximately twice as large at around 38 °C. Again, the highest survival rate was observed in patients with hyperthermia and the lowest in patients with hypothermia. They also observed that ambient temperatures were significantly associated with body temperatures; lower ambient temperatures were associated with hypothermia, and higher outside temperatures were associated with hyperthermia. The potential benefits of active warming of patients with infection were further highlighted by Drewry et al [4], who performed a pilot randomized controlled study evaluating the use of therapeutic hyperthermia in patients with sepsis and observed a significantly better survival in the hyperthermia group; however, there were no differences in the primary results.

We can speculate that higher doses of paracetamol and greater use of active cooling probably decreased body temperatures in the P-N group of patients; however, this was not associated with any improvement during ICU treatment. We observed no differences in the use of noradrenalin, parameters of mechanical ventilation, requirement for renal replacement therapy, or organ dysfunction scores, and we observed a trend toward greater ICU and hospital survival in the P-N group. Our results suggest that there is probably no clinical benefit associated with the treatment of physiological-grade fever. Similarly, Zhang et al [17] and Ye et al [18] performed retrospective studies on patients with fever and sepsis, and they observed no beneficial effects of antipyretic therapy and possible harm associated with the use of external cooling. In a prospective randomized controlled trial, Young et al [19] observed approximately 0.5 °C lower body temperature in patients who received paracetamol (4 g daily) compared to patients who received a placebo. They also observed that patients who received more paracetamol experienced a longer ICU stay if they were nonsurvivors, and shorter ICU stay if they were survivors, which was explained by the probable effect of lower body temperature on clinicians’ perception of the patients’ prognosis. Lower body temperature in the paracetamol group was not associated with any improvement in the ICU course of treatment parameters [19,20].

To our knowledge, there is no other data to compare our results regarding the clearance or persistence of pathogens in tracheal aspirates. In 42.7% (n=50) of our patients (ie, the P-P group of patients), the same pathogen persisted in tracheal aspirates 1 week after initial samples were obtained. There were no differences in causative pathogens between the P-N, P-HAP, and P-P groups of patients, and we observed no significant changes in baseline data between the groups, including the rate of change of empiric antibiotic therapy. There was a trend toward higher temperature in the first 48 hours in the P-N group, compared to the P-HAP and P-P groups, possibly indicating that tolerance of hyperthermia could be beneficial for patients who are intubated with pneumonia and sepsis. There were no significant differences in the course of ICU treatment parameters despite greater use of paracetamol and active cooling, indicating that pharmacological or active cooling has no benefit in patients where fever is part of an appropriate response to infection.

There are a number of limitations to our study. We performed a single-center retrospective observational study with all inherent biases associated with this study design. Of the initial 488 patients with pneumonia, 249 were not included because there were no tracheal aspirates 1 week after ICU admission (because of extubation, death, or no clinical need for obtaining samples if the patients were still intubated). Additionally, target temperatures were defined by the treating physicians. However, there are no guidelines regarding target temperatures for this patient population, and we observed no differences in baseline data between the different groups.
Conclusions
To conclude, we observed better survival in patients who developed higher body temperatures in the first 48 hours after admission to the ICU; however, we observed no changes in other treatment parameters. Additionally, we observed greater use of interventions aimed at cooling the patients (use of paracetamol and a trend toward greater use of active cooling) in patients with negative tracheal aspirates 1 week after ICU admission. Additionally, in this group of patients, we observed a trend toward better survival. Our results speak against the use of interventions aimed at the reduction of body temperature and support the strategy of temperature tolerance in patients who are mechanically ventilated with pneumonia and sepsis.

Acknowledgments
The study was supported by an institutional research grant (grant IRP-2022/01-01).

Authors’ Contributions
AM, JF, and ŽK conceptualized the study. AM and PPB developed the methodology. PPB used the software. AM and PPB performed the validation. PPB performed the formal analysis. DG, AG, and NG conducted the investigation and acquired the resources. AM curated the data curation. DG, AG, and NG prepared the original draft. AM, PPB, JF, ŽK, DG, AG, and NG reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

References

https://www.i-jmr.org/2024/1/e52590

Interact J Med Res 2024 | vol. 13 | e52590 | p.197

(page number not for citation purposes)


Abbreviations

APACHE: acute physiology and chronic health evaluation
ICU: intensive care unit
P-HAP: patients with a different pathogen in week 2 tracheal aspirates
P-N: patients with sterile week 2 tracheal aspirates
P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates
SOFA: sequential organ failure assessment
Validation and Refinement of the Sense of Coherence Scale for a French Population: Observational Study

Paul Sebo¹, MSc, MD; Benoît Tudrej², MD, PhD; Augustin Bernard¹, MD; Bruno Delaunay¹, MD; Alexandra Dupuy¹, MD; Claire Malavergne¹, MD; Hubert Maisonneuve¹², MD

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

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

Abstract

Background: Salutogenesis focuses on understanding the factors that contribute to positive health outcomes. At the core of the model lies the sense of coherence (SOC), which plays a crucial role in promoting well-being and resilience.

Objective: Using the validscale Stata command, we aimed to assess the psychometric properties of the French version of the 3-dimension 13-item SOC questionnaire (SOC-13), encompassing the comprehensibility, manageability, and meaningfulness dimensions. We also aimed to determine if a refined scale, assessed through this method, exhibits superior psychometric properties compared to the SOC-13.

Methods: A sample of 880 consecutive primary care patients recruited from 35 French practices were asked to complete the SOC-13. We tested for internal consistency and scalability using the Cronbach α and Loevinger H coefficients, respectively, and we tested for construct validity using confirmatory factor analysis and goodness-of-fit indices (root mean square error of approximation [RMSEA] and comparative fit index [CFI]).

Results: Of the 880 eligible patients, 804 (91.4%) agreed to participate (n=527, 65.6% women; median age 51 years). Cronbach α and Loevinger H coefficients for the SOC-13 were all <0.70 and <0.30, respectively, indicating poor internal consistency and poor scalability (0.64 and 0.29 for comprehensibility, 0.56 and 0.26 for manageability, and 0.46 and 0.17 for meaningfulness, respectively). The RMSEA and CFI were >0.06 (0.09) and <0.90 (0.83), respectively, indicating a poor fit. By contrast, the psychometric properties of a unidimensional 8-item version of the SOC questionnaire (SOC-8) were excellent (Cronbach α=0.82, Loevinger H=0.38, RMSEA=0.05, and CFI=0.97).

Conclusions: The psychometric properties of the 3-dimension SOC-13 were poor, unlike the unidimensional SOC-8. A questionnaire built only with these 8 items could be a good candidate to measure the SOC. However, further validation studies are needed before recommending its use in research.


KEYWORDS

French; sense of coherence; salutogenesis; SOC; Sense of Coherence scale; validation; validscale; well-being; promoting; resilience; reliability; primary care patients; manageability

Introduction

Salutogenesis, a concept developed by Aaron Antonovsky, represents a paradigm shift in health research as an approach focusing on understanding the factors contributing to positive health outcomes rather than merely concentrating on disease prevention [1,2]. At the core of Antonovsky’s salutogenic model lies the concept of the sense of coherence (SOC), a multifaceted concept reflecting an individual’s capacity to comprehend, manage, and find meaning in the world around them, influencing
their ability to cope with stressors and maintain positive well-being [3]. The three interrelated dimensions of SOC include comprehensibility (perceiving the world as ordered/predictable), manageability (belief in coping effectively with stressors), and meaningfulness (finding purpose/motivation in life). A strong SOC fosters a cognitive orientation that enables individuals to perceive their environment as structured/predictable, facilitating a greater understanding of the challenges they encounter. Moreover, the belief in one’s ability to manage stressors effectively empowers individuals to approach difficulties with confidence/resilience. The sense of meaningfulness, derived from finding purpose/motivation in life, further contributes to an individual’s adaptive capacity in the face of stressors. The SOC theory also introduces a unidimensional model that provides a consolidated measure, aiding in a quicker clinical assessment of an individual’s overall SOC. The choice between the 3-dimension and unidimensional models depends on assessment goals and the required depth of information.

The SOC theory has gained considerable attention in health research, with numerous studies exploring its applicability/implications [3,4]. Researchers typically use the 13-item questionnaire (SOC-13) to assess an individual’s level of coherence and its association with various health-related outcomes [4,5]. The questionnaire has been translated into several languages, including French. However, to our knowledge, the French version did not go through validation procedures. A relatively old population-based study evaluated a French version of the SOC-13 scale modified by the authors [6]. This questionnaire (not provided in their article) showed satisfactory internal consistency but only for the unidimensional model, whereas the validity was insufficient.

Given the lack of information on the validity/reliability of the French SOC-13, we aimed to assess its psychometric properties in primary care patients, ensuring its appropriateness/effectiveness for assessing SOC in French-speaking patients. If the psychometric properties of this scale were found to be insufficient, a secondary objective was to develop an alternative version that would be more valid/reliable than the SOC-13. By selecting primary care patients as the target, we explored the SOC concept in a real-world, patient-centered setting, recognizing implications for interventions and the broader relevance to salutogenesis.

Methods

Study Setting

This observational study was performed with primary care patients in France during 2023. We used a professional register of primary care physicians in the Rhône-Alpes region of France and randomly selected 200 physicians using computer-generated random numbers. Five research assistants contacted each selected physician via email until the target number of physicians (n=35) was attained. In case of refusal or no response after 3 reminders, the next practice on the list was contacted. A sample of 880 consecutive patients recruited from these practices (20-25 patients per practice) were asked to complete the French SOC-13 in the waiting room. Eligible participants were nonurgent, French-speaking, adult patients capable of understanding the study.

Ethical Considerations

The study was approved by the Research Ethics Committee of the University College of General Practice (Lyon) (project ID IRB 2023-01-03-01). Informed consent was obtained from all participants and their ability to opt out was ensured. Privacy/confidentiality were maintained through anonymized data.

SOC-13 Scale

The SOC-13 questionnaire has three components: items 2, 6, 8, 9, and 11 are related to comprehensibility; items 3, 5, 10, and 13 are related to manageability; and items 1, 4, 7, and 12 are related to meaningfulness. The questions are rated on a 7-point Likert scale so that the total score ranges from 13 to 91. The coding for items 1, 2, 3, and 7 is reversed. We summarized the 3 subscores and the total score using the median (IQR).

Validation of the French Version of the 3-Dimension SOC-13 and Development of the Unidimensional 8-Item SOC Questionnaire

We used the validscale command [7] in Stata to assess the psychometric properties of the SOC-13 using classical test theory [8]. We assessed both the 3-dimension and unidimensional models with this approach. We tested for internal consistency and scalability using the Cronbach α and Loevinger H coefficients, respectively. A minimum value of 0.70 for Cronbach α and 0.30 for Loevinger H were considered acceptable [9,10]. We tested for construct validity using confirmatory factor analysis and goodness-of-fit indices. To assess the adequacy of the statistical model, we used the root mean square error of approximation (RMSEA) and the comparative fit index (CFI). These indices evaluate the agreement between the observed and expected data according to the specified model. An RMSEA<0.06 and a CFI>0.90 are generally considered to indicate that the model is a good fit [11]. We used the convdiv option to assess convergent/divergent validities through examination of a correlation matrix [7].

We also developed a shorter questionnaire in French that is potentially more reliable/valid and easier to use in primary care than the SOC-13. We removed all problematic items from the SOC-13 by examining the Cronbach α values obtained for each removed item, while keeping at least 2 questions per dimension. This questionnaire consisted of 8 items (items 6, 8, 9, and 11 for comprehensibility; items 10 and 13 for manageability; and items 4 and 12 for meaningfulness).

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

A total of 804 participants agreed to take part in the study (participation rate=91.4%), 65.6% of whom were women (n=527). The median age of the participants was 51 (IQR 30, range 20-93) years. Depending on the item, between 787 and 793 participants responded to the SOC-13 questions. The median score was 23 (IQR 8, range 5-35) for comprehensibility, 18 (IQR 6, range 4-28) for manageability, 21 (IQR 6, range 8-28) for meaningfulness, and 62 (IQR 16, range 30-89) for the total score.

Internal consistency and scalability were not sufficient for the 3-dimension model. Cronbach \( \alpha \) and Loevinger H coefficients were all <0.70 and <0.30, respectively (0.64 and 0.29 for comprehensibility, 0.46 and 0.17 for meaningfulness, respectively). Table 1 shows the proportion of missing data for each item, the distribution of item responses, and the Loevinger H and Cronbach \( \alpha \) coefficients obtained by omitting each item. Loevinger H coefficients were <0.30 for 9 of the 13 items.

The confirmatory factor analysis, goodness-of-fit indices, and correlation matrix are shown in Table 2. The RMSEA and CFI were 0.09 and 0.83, respectively, indicating a poor fit. Only 4 items had a correlation coefficient with the score of their own dimension >0.40 (indicating lack of convergent validity) and only 5 items had a correlation coefficient with the score of their own dimension greater than those computed with other scores (indicating lack of divergent validity).

Table 1. Distribution of item responses, internal consistency, and scalability of the French versions of the 3-dimension 13-item and unidimensional 8-item sense of coherence (SOC) scales.

<table>
<thead>
<tr>
<th>Scales and items</th>
<th>Missing data, %</th>
<th>Patients, n</th>
<th>Response category, %</th>
<th>Cronbach ( \alpha ) (^a)</th>
<th>Loevinger H</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOC-13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension 1: comprehensibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td>1.37</td>
<td>793</td>
<td>8.45 19.29 30.77 22.32 10.09 6.43 2.65</td>
<td>0.71</td>
<td>0.10</td>
</tr>
<tr>
<td>Item 6</td>
<td>1.37</td>
<td>793</td>
<td>3.28 7.44 10.34 16.27 19.17 27.49 16.02</td>
<td>0.53</td>
<td>0.35</td>
</tr>
<tr>
<td>Item 8</td>
<td>1.37</td>
<td>793</td>
<td>3.28 7.06 12.11 19.55 15.51 25.35 17.15</td>
<td>0.51</td>
<td>0.37</td>
</tr>
<tr>
<td>Item 9</td>
<td>2.11</td>
<td>787</td>
<td>2.03 5.84 11.56 18.17 14.49 25.54 22.36</td>
<td>0.55</td>
<td>0.33</td>
</tr>
<tr>
<td>Item 11</td>
<td>2.11</td>
<td>787</td>
<td>3.30 8.51 12.33 22.24 19.44 23.38 10.80</td>
<td>0.61</td>
<td>0.26</td>
</tr>
<tr>
<td>Dimension 2: manageability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 3</td>
<td>1.74</td>
<td>790</td>
<td>8.73 16.58 29.49 18.10 12.66 11.27 3.16</td>
<td>0.57</td>
<td>0.20</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.37</td>
<td>793</td>
<td>3.40 5.80 11.98 20.18 17.40 24.46 16.77</td>
<td>0.50</td>
<td>0.25</td>
</tr>
<tr>
<td>Item 10</td>
<td>1.87</td>
<td>789</td>
<td>3.17 11.15 15.72 17.49 13.94 20.66 17.87</td>
<td>0.43</td>
<td>0.30</td>
</tr>
<tr>
<td>Item 13</td>
<td>2.11</td>
<td>787</td>
<td>2.54 6.35 10.67 14.23 16.90 31.77 17.53</td>
<td>0.46</td>
<td>0.29</td>
</tr>
<tr>
<td>Dimension 3: meaningfulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>1.37</td>
<td>793</td>
<td>5.04 5.42 11.73 14.75 14.00 25.98 23.08</td>
<td>0.55</td>
<td>0.06</td>
</tr>
<tr>
<td>Item 4</td>
<td>1.37</td>
<td>793</td>
<td>2.02 1.77 2.27 13.11 20.81 33.80 26.23</td>
<td>0.31</td>
<td>0.22</td>
</tr>
<tr>
<td>Item 7</td>
<td>1.49</td>
<td>792</td>
<td>1.39 4.17 6.82 20.45 22.35 30.56 14.27</td>
<td>0.34</td>
<td>0.21</td>
</tr>
<tr>
<td>Item 12</td>
<td>1.99</td>
<td>788</td>
<td>4.31 5.46 9.14 13.07 19.04 32.11 16.88</td>
<td>0.32</td>
<td>0.21</td>
</tr>
<tr>
<td>SOC-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td>1.37</td>
<td>793</td>
<td>3.28 7.44 10.34 16.27 19.17 27.49 16.02</td>
<td>0.79</td>
<td>0.40</td>
</tr>
<tr>
<td>Item 8</td>
<td>1.37</td>
<td>793</td>
<td>3.28 7.06 12.11 19.55 15.51 25.35 17.15</td>
<td>0.78</td>
<td>0.43</td>
</tr>
<tr>
<td>Item 9</td>
<td>2.11</td>
<td>787</td>
<td>2.03 5.84 11.56 18.17 14.49 25.54 22.36</td>
<td>0.80</td>
<td>0.37</td>
</tr>
<tr>
<td>Item 11</td>
<td>2.11</td>
<td>787</td>
<td>3.30 8.51 12.33 22.24 19.44 23.38 10.80</td>
<td>0.81</td>
<td>0.34</td>
</tr>
<tr>
<td>Item 10</td>
<td>1.87</td>
<td>789</td>
<td>3.17 11.15 15.72 17.49 13.94 20.66 17.87</td>
<td>0.79</td>
<td>0.41</td>
</tr>
<tr>
<td>Item 13</td>
<td>2.11</td>
<td>787</td>
<td>2.54 6.35 10.67 14.23 16.90 31.77 17.53</td>
<td>0.80</td>
<td>0.38</td>
</tr>
<tr>
<td>Item 4</td>
<td>1.37</td>
<td>793</td>
<td>2.02 1.77 2.27 13.11 20.81 33.80 26.23</td>
<td>0.81</td>
<td>0.33</td>
</tr>
<tr>
<td>Item 12</td>
<td>1.99</td>
<td>788</td>
<td>4.31 5.46 9.14 13.07 19.04 32.11 16.88</td>
<td>0.80</td>
<td>0.35</td>
</tr>
</tbody>
</table>

\(^a\)Cronbach \( \alpha \) is calculated if the item is removed; for example, if item 2 of SOC-13 were removed, the Cronbach \( \alpha \) for comprehensibility (ie, Dimension 1) would increase from 0.64 to 0.71.
## Table 2. Confirmatory factor analysis for the French versions of the 3-dimension 13-item and unidimensional 8-item sense of coherence (SOC) scales, and the correlation matrix for convergent and divergent validity for the French version of the 3-dimension scale.

<table>
<thead>
<tr>
<th>Scales and items</th>
<th>Factor loading (SE)</th>
<th>Intercept (SE)</th>
<th>Error variance</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dimension 1</td>
</tr>
<tr>
<td><strong>SOC-13</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dimension 1: comprehensibility (variance=0.06)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td>1.00</td>
<td>3.37 (0.05)</td>
<td>1.95</td>
<td>0.124</td>
</tr>
<tr>
<td>Item 6</td>
<td>4.29 (0.95)</td>
<td>4.87 (0.06)</td>
<td>1.49</td>
<td>0.520</td>
</tr>
<tr>
<td>Item 8</td>
<td>4.70 (1.04)</td>
<td>4.82 (0.06)</td>
<td>1.32</td>
<td>0.562</td>
</tr>
<tr>
<td>Item 9</td>
<td>3.71 (0.84)</td>
<td>5.03 (0.06)</td>
<td>1.76</td>
<td>0.476</td>
</tr>
<tr>
<td>Item 11</td>
<td>3.28 (0.75)</td>
<td>4.59 (0.06)</td>
<td>1.80</td>
<td>0.361</td>
</tr>
<tr>
<td><strong>Dimension 2: manageability (variance=0.17)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 3</td>
<td>1.00</td>
<td>3.56 (0.06)</td>
<td>2.23</td>
<td>0.340</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.77 (0.28)</td>
<td>4.83 (0.06)</td>
<td>2.10</td>
<td>0.397</td>
</tr>
<tr>
<td>Item 10</td>
<td>2.73 (0.40)</td>
<td>4.62 (0.06)</td>
<td>1.80</td>
<td>0.566</td>
</tr>
<tr>
<td>Item 13</td>
<td>2.29 (0.34)</td>
<td>5.03 (0.06)</td>
<td>1.65</td>
<td>0.516</td>
</tr>
<tr>
<td><strong>Dimension 3: meaningfulness (variance=0.02)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>1.00</td>
<td>4.97 (0.06)</td>
<td>3.06</td>
<td>0.047</td>
</tr>
<tr>
<td>Item 4</td>
<td>5.72 (3.13)</td>
<td>5.55 (0.05)</td>
<td>1.17</td>
<td>0.383</td>
</tr>
<tr>
<td>Item 7</td>
<td>4.27 (2.34)</td>
<td>5.07 (0.05)</td>
<td>1.62</td>
<td>0.289</td>
</tr>
<tr>
<td>Item 12</td>
<td>7.51 (4.10)</td>
<td>5.00 (0.06)</td>
<td>1.63</td>
<td>0.430</td>
</tr>
<tr>
<td><strong>SOC-8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(variance=1.13)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td>1.00</td>
<td>4.87 (0.06)</td>
<td>1.54</td>
<td>—</td>
</tr>
<tr>
<td>Item 8</td>
<td>1.12 (0.07)</td>
<td>4.82 (0.06)</td>
<td>1.32</td>
<td>—</td>
</tr>
<tr>
<td>Item 9</td>
<td>0.91 (0.07)</td>
<td>5.03 (0.06)</td>
<td>1.71</td>
<td>—</td>
</tr>
<tr>
<td>Item 11</td>
<td>0.79 (0.06)</td>
<td>4.58 (0.06)</td>
<td>1.80</td>
<td>—</td>
</tr>
<tr>
<td>Item 10</td>
<td>1.09 (0.07)</td>
<td>4.62 (0.06)</td>
<td>1.73</td>
<td>—</td>
</tr>
<tr>
<td>Item 13</td>
<td>0.91 (0.06)</td>
<td>5.02 (0.06)</td>
<td>1.61</td>
<td>—</td>
</tr>
<tr>
<td>Item 4</td>
<td>0.63 (0.05)</td>
<td>5.55 (0.05)</td>
<td>1.31</td>
<td>—</td>
</tr>
<tr>
<td>Item 12</td>
<td>0.84 (0.07)</td>
<td>5.01 (0.06)</td>
<td>1.87</td>
<td>—</td>
</tr>
</tbody>
</table>

aSOC-13 scale: $\chi^2_{62}=436.36$, $\chi^2/d.f.=7.0$, root mean square error of approximation=0.088, standardized root mean square residual=0.067, comparative fit index=0.831; convergent validity: 4/13 items (30.8%) have a correlation coefficient with the score of their own dimension greater than 0.400; divergent validity: 5/13 items (38.5%) have a correlation coefficient with the score of their own dimension greater than those computed with other scores.

bUnidimensional SOC-8 scale: $\chi^2_{20}=61.90$, $\chi^2/d.f.=3.1$, root mean square error of approximation=0.052, standardized root mean square residual=0.030, comparative fit index=0.973.

cNot applicable; the dimensions are only relevant to the SOC-13 scale.

The results were similar for the unidimensional SOC-13, except that Cronbach $\alpha$ was higher than that found for the 3-dimension model (Cronbach $\alpha=0.79$, Loevinger H=0.24, RMSEA=0.09, CFI=0.82). By contrast, the psychometric properties of the unidimensional SOC-8 were excellent, as shown in Tables 1 and 2 (Cronbach $\alpha=0.82$, Loevinger H=0.38, RMSEA=0.05, CFI=0.97). The median score was 41 (IQR 13, range 8-56). The 3-dimension SOC-8 produced similar results, except that the Cronbach $\alpha$ values were lower (0.71 for comprehensibility, 0.55 for manageability, and 0.50 for meaningfulness; Loevinger H=0.39, 0.39, and 0.36, respectively; RMSEA=0.06; CFI=0.98).

### Discussion

We assessed the psychometric properties of the SOC scale within the framework of classical test theory. In a French primary care patient population, the validity/reliability of the French version of the 3-dimension and unidimensional SOC-13 scale were poor. By contrast, the psychometric properties of the unidimensional SOC-8 were excellent. The properties of the 3-dimension SOC-8 scale were not better than those of the unidimensional model.
Despite the lack of validation studies (the general population study published in 2001 was based on a version of the questionnaire modified by the authors [6]), the SOC-13 has already been used in several studies in French-speaking populations, including in France [13] and Belgium [14]. However, our study findings indicate that the French questionnaire lacks validity and reliability, possibly influenced by language-specific nuances affecting the scale’s psychometric properties.

Researchers interested in assessing the SOC could perhaps use the unidimensional SOC-8 in the future, which showed excellent psychometric properties in our study, although further validation studies are still needed. All three dimensions (comprehensibility, meaningfulness, and manageability) are represented in the 8 questions of the SOC-8. This confirms that they adequately represent the SOC in reality. The difference in psychometric properties between the SOC-8 and the SOC-13 can be explained by the fact that the 5 items excluded in the SOC-8 are perhaps less clear in French and could potentially lead to different interpretations among respondents.

Our study has several limitations. As this study was limited to patients in France, it raises questions about generalizability to other French-speaking populations. Additionally, reproducibility was not assessed and external validation of the SOC-8 is crucial before widespread adoption.

In conclusion, our study suggests that the psychometric properties of the French version of the 3-dimension SOC-13 are poor, unlike the unidimensional SOC-8. A questionnaire built only with these 8 items could be a good candidate to measure SOC. However, further validation studies are needed before recommending its use in research.

Acknowledgments
The authors would like to sincerely thank all the primary care physicians and patients who collaborated on this project. Special thanks go to Amir Moussa for his involvement in this project.

Data Availability
The data associated with this article are available in the Open Science Framework [15].

Authors' Contributions
BT and HM were responsible for study conceptualization. AB, BD, AD, and CM were responsible for project administration. PS performed the formal analysis and wrote the original draft of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The French versions of the SOC-13 and the SOC-8.

References

Abbreviations

- **CFI**: comparative fit index
- **RMSEA**: root mean square error of approximation
- **SOC**: sense of coherence
- **SOC-8**: 8-item sense of coherence scale
- **SOC-13**: 13-item sense of coherence scale

©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/), 16.07.2024. 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.
Review

Gamification and Oral Health in Children and Adolescents: Scoping Review

Rui Moreira1,2*, MD; Augusta Silveira3,4,5*, MSc, DMD, PhD; Teresa Sequeira3,4,5*, PhD; Nuno Durão2*, MD; Jessica Lourenço2*, MD; Inês Cascais6*, MD; Rita Maria Cabral2*, MD; Tiago Taveira Gomes3,7*, MSc, ME, MD, PhD

1School of Medicine and Biomedical Sciences, Instituto de Ciencias Biomédicas Abel Salazar, University of Porto, Porto, Portugal
2Department of Stomatology and Oral Maxillofacial Surgery, University Hospital Center of Santo António, Porto, Portugal
3Faculty of Health Sciences, Fernando Pessoa University, University of Porto, Porto, Portugal
4Institute for Research, Innovation and Development, Fernando Pessoa Foundation, University of Porto, Porto, Portugal
5Center for Innovative Biomedicine and Biotechnology, University of Coimbra, Coimbra, Portugal
6Paediatrics Department, Maternal & Child Center of the North, University Hospital Center of Santo António, Porto, Portugal
7Department of Community Medicine, Information and Decision in Health, Medical School of Porto, University of Porto, Porto, Portugal

*all authors contributed equally

Corresponding Author:
Rui Moreira, MD
School of Medicine and Biomedical Sciences
Instituto de Ciencias Biomédicas Abel Salazar
University of Porto
Rua de Jorge Viterbo Ferreira, 228
Porto, 4050-313
Portugal
Phone: 351 914180479
Email: ruidcmoreira@gmail.com

Abstract

Background: Oral health is a determinant of overall well-being and quality of life. Individual behaviors, such as oral hygiene and dietary habits, play a central role in oral health. Motivation is a crucial factor in promoting behavior change, and gamification offers a means to boost health-related knowledge and encourage positive health behaviors.

Objective: This study aims to evaluate the impact of gamification and its mechanisms on oral health care of children and adolescents.

Methods: A systematic search covered multiple databases: PubMed/MEDLINE, PsycINFO, the Cochrane Library, ScienceDirect, and LILACS. Gray literature, conference proceedings, and WHOQOL internet resources were considered. Studies from January 2013 to December 2022 were included, except for PubMed/MEDLINE, which was searched until January 2023. A total of 15 studies were selected following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The eligibility criteria were peer-reviewed, full-text, and empirical research related to gamification in oral health care, reports of impact, and oral health care outcomes. The exclusion criteria encompassed duplicate articles; unavailable full texts; nonoriginal articles; and non-digital game-related, non-oral health-related, and protocol studies. Selected studies were scrutinized for gamification mechanisms and outcomes. Two main questions were raised: “Does gamification in oral health care impact oral health?” and “Does oral health care gamification enhance health promotion and literacy?” The PICO (Patient, Intervention, Comparison, Outcome) framework guided the scoping review.

Results: Initially, 617 records were obtained from 5 databases and gray literature sources. After applying exclusion criteria, 15 records were selected. Sample size in the selected studies ranged from 34 to 190 children and adolescents. A substantial portion (11/15, 73%) of the studies discussed oral self-care apps supported by evidence-based oral health. The most clearly defined data in the apps were “brushing time” (11/11, 100%) and “daily amount brushing” (10/11, 91%). Most studies (11/15, 73%) mentioned oral health care behavior change techniques and included “prompt intention formation” (11/26, 42%), “providing instructions” (11/26, 42%), “providing information on the behavior-health link” (10/26, 38%), “providing information on consequences” (9/26, 35%), “modeling or demonstrating behavior” (9/26, 35%), “providing feedback on performance” (8/26, 31%), and “providing contingent rewards” (8/26, 31%). Furthermore, 80% (12/15) of the studies identified game design elements incorporating...
gamification features in oral hygiene applications. The most prevalent gamification features were “ideological incentives” (10/12, 83%) and “goals” (9/16, 56%), which were found in user-specific and challenge categories, respectively.

**Conclusions:** Gamification in oral health care shows potential as an innovative approach to promote positive health behaviors. Most studies reported evidence-based oral health and incorporated oral health care behavior change techniques.


**KEYWORDS**
gamification; mechanisms of gamification; gamification components; intrinsic and extrinsic motivators; oral health care; health behavior; oral health care applications

**Introduction**

People’s health behaviors, such as physical activity, diet, tobacco and alcohol use, recreational drug consumption, and adherence to chronic medications, directly influence their health risks and consequent diseases. To decrease the burden of preventable chronic diseases and enhance well-being in society, it is essential to bring about a change in behavior [1-4]. According to the World Health Organization, oral health is a strong indicator of general health, well-being, and quality of life. The treatment of oral pathologies is expensive and usually not covered under universal health care, accounting for 5% of health care expenditures and 20% of out-of-pocket expenses in wealthy countries [5-11]. Health care systems have limited resources, and making informed decisions based on data collection, focusing on individuals, can provide better health outcomes without incurring additional costs; this approach can deliver better value and at the same time reduce costs [12,13].

Motivation is a core target of a wide range of established behavior change techniques [1,14-17]. Computer games can be used to increase health-related knowledge and promote desirable health behaviors in children [18]. Games are designed to provide enjoyment, engagement, and satisfaction [1,19-22]. Mobile phones and mobile health technologies can address these issues at low costs [1,23,24]. Mobile devices are useful for delivering health interventions due to their widespread adoption, powerful technical capabilities, and portability [25]. The positive emotional attachment with the user may increase the benefit of health promotion via mobile devices, allowing health interventions to be delivered immediately, anytime, and anywhere [26]. The use of health care apps provides easy access to information and has the potential to improve patient engagement and treatment compliance [27]. Indeed, the number of health care apps available has been growing year after year, with over 200 billion app downloads worldwide from the Apple App Store and Google Play in 2020 [28].

Bohn et al [29] found that educational applications are valuable tools for enhancing patient-provider communication in dental settings. Studies have pointed out that the traditional educational approach, which relies mainly on reading and listening to standardized content, should be replaced with customizable and interactive involvement, using communication tools that are familiar to newer generations [30]. Gamification is a possible response to overcoming the challenges of communication and motivation in health care [31], as it can track individual behaviors and involve users in goal-chasing activities while displaying progress and feedback through personalized information apps [1,32].

Deterding et al [33] define gamification as the integration of game design elements into nongame contexts [34]. This process enriches products, services, and information systems with game design features to positively influence the motivation, productivity, and behavior of users [33,35,36]. Gamified systems commonly use motivational features such as immediate success feedback (reward mechanism), continuous progress feedback, and goal setting. These systems work through interface elements such as point-scoring systems, badges, levels, challenges, competitions, relatedness support, social feedback (engagement loops), recognition, comparison through leaderboards, teams, communication functions, autonomized support through customizable avatars and environments, and narratives providing emotional and value-based rationales for certain activities (customization) [1,34,37,38]. Users receive badges that represent success and can be shared in social networks and displayed in a digital trophy cabinet when new milestones are achieved [35]. Recently, the self-determination theory has become a key framework for health behavior interventions and research [1,39-43]. According to this theory, gamification induces 4 main intrinsic motivators: user satisfaction, conveyance of optimism, provision of meaning [35,44], and facilitation of social interactions [45,46].

Apart from financial aspects, extrinsic motivators systematically activate intrinsic motivators, such as social recognition, support of learning processes, and behavioral change. The evidence suggests that positive outcomes are stronger when gamification is used to target behavioral outcomes [1,23,24,47-50]. However, critics have pointed out the lack of high-quality effect studies on gamification [22,38,51,52]. Nevertheless, clinicians have an opportunity to promote engagement in health promotion through a motivating, fulfilling, and fun activity [53]. Therefore, this study aims to analyze the impact of gamification and its mechanisms on oral health care.

The objective of this scoping review was to analyze the impact of gamification and its mechanisms on oral health care in children and adolescents. We assessed the effectiveness of gamification in promoting changes in oral health behavior and enhancing oral health care outcomes for its users. Furthermore, we investigated the role of gamification in oral health care, including the integration of evidence-based oral health care concepts and gamification design elements in application design.
Methods

Objective
This scoping review aims to analyze the impact and efficiency of gamification mechanisms on oral health care, with a focus on promoting user engagement to expand oral health literacy and support oral health care policies, in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [54,55].

Scoping Review
This review used a systematic methodology to identify gamification on oral health care apps, evaluate their features, identify their gamification mechanisms, and follow their outcomes.

We did not apply any restrictions related to population samples or oral health care systems, but we excluded contexts outside oral health.

The following main questions guided our analysis:
- Does gamification in oral health care impact oral health?
- Does oral health care gamification enhance health promotion and literacy?

On the basis of these questions, we searched several domains, including oral health care outcomes, gamification design elements, evidence-based oral health care concepts used in app development, and oral health behavior changes related to gamification. We use the PICO (Patient, Intervention, Comparison, Outcome) framework to elucidate the systematic review questions:
- P (Patient, Population, or Problem): the population under investigation included children and adolescents: mother-preschooler (3-6 years old) dyads and adolescents (younger than 16 years).
- I (Intervention): the intervention under investigation was the gamification strategy and its usefulness for oral health promotion and literacy.
- C (Comparison): gamification was compared with its alternative, traditional learning methods.
- O (Outcome): the outcome assessed was divided into 2 main sections: behavior change techniques and gamification mechanisms.

Search Strategy
The search was conducted across several electronic bibliographic databases, including PubMed/MEDLINE, PsycINFO, the Cochrane Library, ScienceDirect, and LILACS. In addition, grey literature, conference proceedings, and WHOQOL internet resources were assessed. The search strategy included terms related to gamification and oral health care such as gamification, oral health care, policies, games, digital, apps, and outcomes. The authors used a controlled and hierarchically organized vocabulary produced by the National Library of Medicine called the Medical Subject Headings to ensure that the search results accurately reflected the subject content of journal articles as they are published. The search strategy enabled us to identify both published and unpublished studies. All sources were last searched until July 2023, except for PubMed/MEDLINE, which was searched until January 2023. The references to gamification date back to nearly 2010, and articles exploring gamification in oral health were only mentioned or studied in the last decade. There were no language restrictions, and studies published between January 2013 and December 2022 (10 years) were included.


Initially, all types of articles were considered eligible, including systematic reviews, research articles, and prospective and retrospective studies, as long as they met the following criteria: (1) peer-reviewed, (2) full-text papers, (3) empirical research (qualitative and quantitative), (4) explained research methods, (5) gamification as a research subject, (6) effect reported in terms of impact (affect, behavior, and cognition) and user experience, and (7) oral health care outcomes. Criteria 1 to 4 were implemented to ensure focus on high-quality work reporting original research. Criteria 3, 4, and 7 were also included to enable assessment of the quality of evidence. Criterion 5 admitted papers that studied gamification in a broader concept, even if it did not elicit game elements. Criteria 6 and 7 were chosen to assess reported health and well-being outcomes and potential mediators, with user experience being included given its prevalence as an outcome.

However, studies were excluded if they (1) were duplicated; (2) had full text not available; (3) were not the original article; (4) did not refer to a game; (5) were nondigital, such as conventional games like cards or board games; (6) did not concern oral health; and (7) were study protocols without outcomes to report.

Study Quality Assessment, Data Extraction, and Analysis Plan
All searched articles were filtered using broad selection criteria framed as questions:
- Does gamification in oral health care impact oral health?
- Does oral health care gamification enhance health promotion and literacy?

The study selection and data extraction were performed blindly. After the search, all references were imported into a reference management system (Mendeley), and duplicates were removed. The remaining articles’ titles and abstracts were assessed to identify eligible studies. To determine eligibility, three additional questions were asked and answered: (1) Is the topic relevant to the defined scope? (2) Does it meet the inclusion and exclusion criteria? and (3) Is the methodology appropriate? To ensure a comprehensive, transparent, and objective extraction process, a standardized pre piloted form was used to extract data from the included studies. Two reviewers independently extracted the data, and any discrepancies were resolved through discussion with a third author. Additionally, the third reviewer further scrutinized the data to verify the consistency of the extraction...
process and resolved any remaining discrepancies. In case of missing or additional data, other researchers were contacted [56,57].

The eligibility assessment of each full-text paper was conducted by 2 independent raters. In cases where discrepancies arose, they were resolved through discussion and comparison of our evaluations. Articles were excluded from our review when both investigators unanimously concurred on their ineligibility due to inappropriate methodology or results that did not address the key research questions at hand. This rigorous and collaborative approach to eligibility assessment ensured the quality and relevance of the articles included in our study.

Each article was classified as having low, moderate, or high relevance. Articles were deemed highly relevant if they effectively demonstrated an impact on the considered items, whereas moderate relevance was attributed to those that projected such items. Low relevance was assigned to manuscripts that did not present any conclusions or perspectives in these domains.

Data from the scoping literature were extracted into an Excel (Microsoft Corp) data sheet using a support checklist. The data sheet was divided into sections dedicated to a theory, area, concept, theme, or element from the framework of gamification and oral health care, including game mode, population sample, gamification components, behavior change techniques, and outcomes. After synthesizing the data and assessing the quality of the evidence, the writing of the scoping review article began.

Results

Process Selection

Figure 1 shows the PRISMA flowchart outlining the process of record identification, selection, eligibility, and inclusion. Initially, 617 records were retrieved from 5 databases and gray literature sources. After removing duplicates and records that did not meet the inclusion criteria, 15 records were selected for analysis.

![PRISMA flowchart](image-url)
Sample Characteristics

Of the 15 studies available (see Multimedia Appendix 2 [9,18,26,27,30,58-67]), the majority (n=8, 53%) collected data at multiple time points (2 or more) from various points or conditions [18,30,58-62,68]. Of these 8 studies, 7 were blind, randomized control trials [18,30,58,59,61-63]. A content analysis of the applications for behavior change was adopted in 7 studies [9,27,60-66,64,65]. One-time cross-sectional questionnaires were applied in 3 studies [26,66,67].

The sample size in the selected studies ranged from 34 to 190 individuals. All participants were children younger than 13 years, except for 4 studies that also included adults [26,61,62,65]. Mobile apps were the predominant modality used to change oral health care behavior in the studies (n=12), with 3 exceptions based on computer games. One of the games was available on a tablet and DVD for PC [63], whereas the other 2 apps were associated with a toothbrush sensor [58,59].

Oral Hygiene Evidence-Based Categories

The oral self-care applications supported by evidence-based oral health were mentioned in 73% (11/15) of the selected studies. The group categories and archetypes of evidence-based content related to oral hygiene were identified by analyzing the selected studies, as illustrated in Figure 2.

We found that the most clearly defined types of information in the applications were “brushing time” for at least 2 minutes (11/11, 100%) and “daily amount brushing” at least twice a day (10/11, 91%). However, “toothpaste spit out after” brushing was only found in 1 application (1/12, 8%), and “cleaning the tongue” and “toothbrush grip” were never mentioned.

Figure 2. Evidence-based content related to group categories and archetypes of oral hygiene, as recognized in the chosen studies. The y-axis displays the listed group categories and archetypes, whereas the x-axis signifies the frequency of their occurrence within the studies. The accompanying legend elucidates the color-coding system denoting distinct group categories and archetypes. TB: toothbrush; TP: toothpaste.

Behavior Change Techniques

Behavior change techniques for oral health care resulted from mechanisms of gamification, either implicit or explicit. Oral health care behavior change techniques were mentioned in 73% (11/15) of the studies. The data collected from these studies were used to create a behavior change score with 26 items [9]. The studies were analyzed to determine the frequency of behavior change techniques used in oral health applications, as shown in Table 1.

Among the behavior change techniques scrutinized, a set of 7 distinct components emerged as the most prevalently used across the array of surveyed studies. These components, along with their respective frequencies of use, are outlined as follows: “prompt intention formation” (11/26, 42%), “provide instructions” (11/26, 42%), “provide information on behavior-health link” (10/26, 38%), “provide information on consequences” (9/26, 35%), “model or demonstrate behavior” (9/26, 35%), “provide feedback on performance” (8/26, 31%), and “provide contingent rewards” (8/26, 31%).
Table 1. Frequency distribution of behavior change techniques observed in oral health applications identified in the selected studies (N=137).

<table>
<thead>
<tr>
<th>Behavior change techniques</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Provide information behavior health link</td>
<td>10 (7.3)</td>
</tr>
<tr>
<td>2. Provide information on consequences</td>
<td>8 (6.6)</td>
</tr>
<tr>
<td>3. Provide information seeking others’ approval</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td>4. Prompt intention formation</td>
<td>11 (8)</td>
</tr>
<tr>
<td>5. Prompt barrier identification</td>
<td>3 (2.2)</td>
</tr>
<tr>
<td>6. Provide general encouragement</td>
<td>7 (5.1)</td>
</tr>
<tr>
<td>7. Set graded tasks</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>8. Provide instruction</td>
<td>11 (8)</td>
</tr>
<tr>
<td>9. Model or demonstrate behavior</td>
<td>9 (6.6)</td>
</tr>
<tr>
<td>10. Prompt specific goal setting</td>
<td>4 (2.9)</td>
</tr>
<tr>
<td>11. Prompt review of the behavioral goals</td>
<td>5 (3.6)</td>
</tr>
<tr>
<td>12. Prompt self-monitoring behavior</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>13. Provide feedback on performance</td>
<td>8 (5.8)</td>
</tr>
<tr>
<td>14. Provide contingent rewards</td>
<td>8 (5.8)</td>
</tr>
<tr>
<td>15. Teach to use prompts or cues</td>
<td>4 (2.9)</td>
</tr>
<tr>
<td>16. Agree on behavioral contract</td>
<td>7 (5.1)</td>
</tr>
<tr>
<td>17. Prompt practice</td>
<td>7 (5.1)</td>
</tr>
<tr>
<td>18. Use follow-up prompts</td>
<td>4 (2.9)</td>
</tr>
<tr>
<td>19. Provide social comparison</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>20. Plan social support or change</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>21. Prompt identification as a role model</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>22. Prompt self-talk</td>
<td>2 (1.5)</td>
</tr>
<tr>
<td>23. Relapse prevention</td>
<td>4 (2.9)</td>
</tr>
<tr>
<td>24. Stress management</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td>25. Motivational interviewing</td>
<td>5 (3.6)</td>
</tr>
<tr>
<td>26. Time management</td>
<td>7 (5.1)</td>
</tr>
</tbody>
</table>

Gamification Mechanisms

The game design elements based on the gamification features rating criteria for oral hygiene applications were recognized in 12 (80%) of the 15 studies. Table 2 illustrates the 26 gamification features [9] that were considered, along with the number of times each element was identified.

On average, the applications included in the study possessed an average of 10.6 of 31 potential gamification features. Notably, specific game design elements were prevalently used within various categories. Among the system design features, “meaning” (10/36, 28%) stood out prominently. Within the challenges category, “goals” (9/16, 56%) emerged as the most frequently incorporated element. For rewards, “ownership” (9/22, 41%) was notably prevalent. Among social influences, both “collaboration” and “reputation” (both 3/19, 16%) were prominent. Additionally, within the user-specific category, “ideological incentives” (10/12, 83%) exhibited a substantial presence. Interestingly, none of the applications used “badges,” “conforming behavior,” “virtual goods,” or “self-expression” (Table 2).
Table 2. Frequency distribution of game design elements detected in the surveyed studies.

<table>
<thead>
<tr>
<th>Game design elements</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System design (n=36)</strong></td>
<td></td>
</tr>
<tr>
<td>Visual feedback</td>
<td>9 (25)</td>
</tr>
<tr>
<td>Audible feedback</td>
<td>4 (11)</td>
</tr>
<tr>
<td>Reminder</td>
<td>5 (14)</td>
</tr>
<tr>
<td>Meaning</td>
<td>10 (28)</td>
</tr>
<tr>
<td>Integration concepts</td>
<td>3 (8)</td>
</tr>
<tr>
<td>Visually resembling games</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Fantasy</td>
<td>4 (11)</td>
</tr>
<tr>
<td><strong>Challenges (n=16)</strong></td>
<td></td>
</tr>
<tr>
<td>Goals</td>
<td>9 (56)</td>
</tr>
<tr>
<td>Time pressure</td>
<td>4 (25)</td>
</tr>
<tr>
<td>Progressive disclosure</td>
<td>3 (19)</td>
</tr>
<tr>
<td><strong>Rewards (n=22)</strong></td>
<td></td>
</tr>
<tr>
<td>Ownership</td>
<td>9 (41)</td>
</tr>
<tr>
<td>Achievement</td>
<td>5 (23)</td>
</tr>
<tr>
<td>Point system</td>
<td>4 (18)</td>
</tr>
<tr>
<td>Badges</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Bonus</td>
<td>4 (18)</td>
</tr>
<tr>
<td><strong>Social influences (n=19)</strong></td>
<td></td>
</tr>
<tr>
<td>Loss aversion</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Status</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Collaboration</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Reputation</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Competition</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Envy</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Shadowing</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Social facilitation</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Conforming behavior</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Leaderboards</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Altruism</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Virtual goods</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>User specifics (n=12)</strong></td>
<td></td>
</tr>
<tr>
<td>User levels</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Ideological incentives</td>
<td>10 (83)</td>
</tr>
<tr>
<td>Virtual characters</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Self-expression</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

The oral health outcomes related to gamification interventions highlighted the role of gamification in promoting oral health care and literacy. This scoping review also highlights the limitations of currently available oral health care apps and points out the main areas to invest in for the future. A total of 11 (73%) of the 15 articles found positive impacts of using oral health apps, especially in children and adolescents. They facilitate the responsiveness of oral preventive care [66]; improve knowledge in high-risk populations; encourage dietary changes [63]; and promote a reduction of clinical plaque, gingival, and caries
indexes [26]. Additionally, they show a statistically significant improvement in health care indices [58], tooth brushing quality (duration and distribution) [60], and motivation to brush teeth for longer [26] and seem effective in adolescents with fixed orthodontic appliances by self-reported behavior and psychosocial factors [59]. Gamification structures augment oral health literacy, facilitate user alertness for oral health care themes and professional feedback, and engage commitment. A greater improvement in gingival status is commonly reported [27,61,62,69].

The feedback provided by participants showed a higher level of satisfaction in learning about oral health care through games rather than traditional noninteractive methods. Most studies reported a positive impact of gamification, particularly in children and adolescents, who are considered the main target audience of these apps [18,26,30,58-66].

The studied apps contained educational content with evidence-based dentistry and high-quality teaching for oral self-care. Some of these also featured gamification elements and behavior change techniques. The results of the studies demonstrate that these apps have excellent functionality, effectiveness, efficiency, and user satisfaction [9,64,66]. Several studies evaluating multiple oral hygiene apps have found evidence-based content, such as brushing time and daily amount of brushing. Fijačko et al [9], Parker et al [27], and Hotwani et al [64] all reported on these elements.

The health behavior change techniques found in the analyzed apps included prompt intention formation, shaping and demonstrating behavior, providing information about the link between behavior and health consequences, instructions, and contingent rewards [26,30,58,63,66,67]. A set of 7 distinct components emerged as the most prevalently employed across the array of surveyed studies. These components, along with their respective frequencies of use, are outlined as follows: “prompt intention formation,” “provide instructions,” “provide information on behavior-health link,” “provide information on consequences,” “model or demonstrate behavior,” “provide feedback on performance,” and “provide contingent rewards.” These components collectively represent the core elements of behavior change techniques that were consistently integrated into the analyzed oral health applications, aiming to enhance engagement and promote positive behavior change.

Regarding game design elements, these applications emphasized feedback, goal attainment, sense of ownership, and ideological incentives [27,61,65-67]. Parker et al [27] identified some recurring game design elements among 20 apps analyzed, such as knowledge provision, self-monitoring of frequency, and duration of toothbrushing. Hotwani et al [64] found that information provision, goal setting, feedback, progressive disclosure, and time pressure were frequently used in the 6 apps evaluated. Fijačko et al [9] analyzed 17 apps and identified time pressure, digital characters, and fantasy as key game design elements.

Comparison With Prior Work

Delivering trustworthy information to users is essential for promoting healthy habits. Health care apps should undergo validation by health care institutions and professionals before becoming public to ensure their accuracy and reliability. However, there is a risk of users becoming overly dependent, potentially compromising the need for regular appointments with oral health care professionals in real life [27,70].

Considering evidence-based oral health care, most apps emphasize brushing for at least 2 minutes and twice a day. Although this is an important core of oral hygiene recommendations, there is still room for improvement. To achieve holistic oral care, it would be advisable to incorporate other aspects such as oral hygiene techniques, the use of devices, dietary advice, sugar intake control, guidance on early childhood caries, baby oral hygiene, the effects of fluoride, the use of fluoride toothpaste, toothbrushing training videos, and regular dental visits. The development of apps should be based on theoretical models when designing educational content, and the accuracy of the content should be a priority to bring about real behavior change [61,64,70]. Sharif and Alkadhim [65] went beyond the basics and included interdental cleaning, spitting out after brushing, avoiding mouth rinsing after brushing, characteristics of the toothbrush, and the quantity of fluoride content in toothpaste and mouthwash. Other advisable strategies include reporting about others’ approval, social interactions with oral health professionals and other application users, identifying barriers to oral hygiene and potential overcoming strategies, providing encouragement, setting graded tasks and goals, displaying tracked data and objectives, feedback on performance, setting a behavioral contract with oral health professionals, social comparison, and social support [1,9,27,35,63,65,71].

Patient adherence to a smartphone app is more effective because of the ability to customize reminders and prompts, constant accessibility, adjustability to the user, ability to provide tailored feedback, widespread use, and interactive features [63,72,73].

Strengths and Limitations

Despite the proven efficiency of gamification in health care promotion and prevention, it remains an unexplored territory in oral health, mainly applied to specific educational purposes and oral health promotion [74]. This scoping review highlights the limitations of currently available oral health care applications and points out the main areas to invest in for the future. Two major limitations were found in this study. The first is the limited availability of articles related to the main topics. Gamification, within the context of health care, is not a recent concept. However, its quantitative assessment, particularly in the field of oral health, remains relatively uncommon. To address such limitation and ensure a comprehensive review, we diligently accessed and explored 5 databases, along with incorporating gray literature sources. This approach was essential to include as many relevant studies as possible, aligning with our predefined inclusion criteria. Considering the diverse nature of our search strategy, we believe that we have made every effort to provide a reliable representation of the existing literature in the field of gamification in oral health care.

The second limitation is the heterogeneous studies with varying focus and dispersing attention. The diversity of approaches makes a rigorous comparison more challenging.
Taking into account the multiple aspects involved in gamification strategies and by transparently outlining these parameters, we believe that our work can serve as a valuable reference for future researchers seeking to design studies that address and overcome these challenges. Our hope is that this will pave the way for a more effective understanding of the underlying mechanisms behind the implementation of gamification in oral health context.

**Future Directions**

Future studies could focus on other age groups as well, such as the study by Zolfaghari et al [61], which developed applications for mothers of children, improving their oral health literacy and practice and promoting plaque control in children within just 1 month of use [61,74,75].

There is potential for future optimization of key gamification features, such as badges, encouragement of correct behavior, digital goods, self-expression, reminders, fantasy themes, time pressure, disclosure of progress, achievements, points systems, bonuses, loss aversion, status, collaboration, reputation, competition, shadowing, social facilitation, leaderboards, altruism, user levels, and digital characters [26,27,30,66,71,76].

**Conclusions**

Gamification in oral health care does have an impact; it enhances oral health promotion and literacy. It represents a potential new approach for oral health care providers to change people’s oral health behavior. The most frequent game design mechanisms adopted were meaning, ideological incentives, feedback, goals, and ownership. Some authors have highlighted several factors for gamification success, including engagement strategy, applications aesthetics, evidence-based information content, behavioral change taxonomies, attention to psychological needs, evaluation, validation, quality assessment, and professional regulation standards for oral health care applications. More studies are needed to better understand the clinical, psychological, and social processes involved in selecting the most efficient gamification mechanisms. The process of mobile health in oral health care is in the initial stage, but gamification is crucial for improving individual health-related practices.

**Acknowledgments**

The authors would like to express their gratitude to Professor Tiago Taveira-Gomes for his mentorship, review of the manuscript, and unwavering support.

**Conflicts of Interest**

None declared.

---

**References**


Abbreviations

PICO: Patient, Intervention, Comparison, Outcome
Gamification and Oral Health in Children and Adolescents: Scoping Review

Please cite as:
Moreira R, Silveira A, Sequeira T, Durão N, Lourenço J, Cascais I, Cabral RM, Taveira Gomes T

URL: https://www.i-jmr.org/2024/1/e35132
doi:10.2196/35132
PMID: 38573750

©Rui Moreira, Augusta Silveira, Teresa Sequeira, Nuno Durão, Jessica Lourenço, Inês Cascais, Rita Maria Cabral, Tiago Taveira Gomes. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 04.04.2024. 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.
Evaluation of the Accuracy, Credibility, and Readability of Statin-Related Websites: Cross-Sectional Study

Eunice Ling¹, BAppSc, MD; Domenico de Pieri¹*, MD; Evenne Loh¹**, MD; Karen M Scott¹, BEd, MA, PhD; Stephen C H Li², MBBS, MSci, MPH, MBA, MAACB; Heather J Medbury¹,³, BSc, PhD

¹Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia
²Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, Australia
³Vascular Biology Research Centre, Surgery, Westmead Hospital, Westmead, Australia
* these authors contributed equally

Corresponding Author:
Heather J Medbury, BSc, PhD
Vascular Biology Research Centre
Surgery
Westmead Hospital
REN Building
Westmead, 2145
Australia
Phone: 61 8890 3668
Email: heather.medbury@sydney.edu.au

Abstract

Background: Cardiovascular disease (CVD) represents the greatest burden of mortality worldwide, and statins are the most commonly prescribed drug in its management. A wealth of information pertaining to statins and their side effects is on the internet; however, to date, no assessment of the accuracy, credibility, and readability of this information has been undertaken.

Objective: This study aimed to evaluate the quality (accuracy, credibility, and readability) of websites likely to be visited by the general public undertaking a Google search of the side effects and use of statin medications.

Methods: Following a Google web search, we reviewed the top 20 consumer-focused websites with statin information. Website accuracy, credibility, and readability were assessed based on website category (commercial, not-for-profit, and media), website rank, and the presence or absence of the Health on the Net Code of Conduct (HONcode) seal. Accuracy and credibility were assessed following the development of checklists (with 20 and 13 items, respectively). Readability was assessed using the Simple Measure of Gobbledegook scores.

Results: Overall, the accuracy score was low (mean 14.35 out of 20). While side effects were comprehensively covered by 18 websites, there was little information about statin use in primary and secondary prevention. None of the websites met all criteria on the credibility checklist (mean 7.8 out of 13). The median Simple Measure of Gobbledegook score was 9.65 (IQR 8.825-10.85), with none of the websites meeting the recommended reading grade of 6, even the media websites. A website bearing the HONcode seal did not mean that the website was more comprehensive or readable.

Conclusions: The quality of statin-related websites tended to be poor. Although the information contained was accurate, it was not comprehensive and was presented at a reading level that was too difficult for an average reader to fully comprehend. As such, consumers risk being uninformed about this pharmacotherapy.


KEYWORDS
statins; consumer health information; readability; credibility; accuracy; digital health, health information seeking; cardiovascular; mortality; management; pharmacotherapy; risk; medication
Introduction

Background
Cardiovascular diseases (CVDs) are the primary cause of death globally, with an estimated 17.9 million people dying of CVDs in 2021. This represents 31% of all global deaths. Of these deaths, 85% are due to heart attack and stroke, whose most common etiology is atherosclerosis [1]—the development of fatty plaque within artery walls. A key pharmacological treatment for atherosclerosis is statin therapy. It has a role in the primary and secondary prevention of vascular events, with a lowering of low-density lipoprotein cholesterol leading by 2 mmol/L, typically giving a 10% absolute benefit (the reduction in the probability of an event’s occurrence within a population receiving treatment) for those diagnosed with vascular disease and a 5% absolute benefit for those with risk factors yet without having experienced a vascular event [2]. This creates issues when we consider that patients may be biochemically abnormal (with hypercholesterolemia) but asymptomatic. Such patients may doubt the use of the prescribed statin therapy as they determine the cost-benefit balance between tangible adverse effects and theoretical benefits. This may prompt the consultation of alternative sources of knowledge to aid decision-making. In this era of shared decision-making, where patients participate in the medical decisions that affect their health [3], it is essential that the information they access is high quality and easily understood.

Use of the Internet for Health Information Seeking
In this milieu, the internet has risen as a key source of health-related information, with 79% of adults seeking web-based health information in America and 79% to 86% in China, the Philippines, Hong Kong, Indonesia, and Vietnam [4,5]. Comparatively, seeking web-based health information is as popular as playing games or downloading music from the web [6]. Notably, the COVID-19 pandemic has presented unprecedented challenges, catapulting society further into a future dependence on telehealth and internet-assisted health care [7,8]. As such, traditional in-clinic and leaflet modes of health information delivery are being supplemented, and in some cases supplanted, by internet searches. With this dramatic change in the terrain upon which patients and their families are attaining information, it is crucial to determine the quality of web-based health information put forth to them.

Health literacy is defined as “the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” [9]. It requires a complex group of skills such as reading, listening, analyzing, and decision-making, as well as the ability to apply the aforementioned skills to health situations [10]. Those with poor health literacy are vulnerable to undertaking unnecessary tests and treatments or, conversely, refusing beneficial tests and treatments. In part, they may be misled into assessing the quality of web-based health information based on its search result ranking, image quality, celebrity endorsement, and website authorship rather than relying on the criteria of established quality guidelines [11,12].
we determined the relationship between the accuracy, credibility, and readability of the websites found.

In selecting websites on statins to analyze, we aimed to emulate a typical consumer’s search for web-based health information. A web search was conducted using Google, in keeping with evidence that it accounts for more than half of all web traffic [28-30] and is an increasingly preferred search engine by the general public [31]: 91% of American adults using the internet use a search engine, and of those, 83% use Google more often than other search engines [4,32]. To conduct the search, location filters, user information, search history, cached data, and cookies were disabled, and sponsored results were excluded to avoid inadvertent search bias. The search terms used were “statin” and “statin side effects,” following the advice of our lipidologist coauthor (SL) that the generic term “statin,” rather than specific medication names, was commonly used in discussion with patients in clinical practice and that “statin side effects” were a key concern of patients.

The first 20 ranked websites in the Google search results page were analyzed (after removing any duplication from search results of the 2 search terms). We did not identify further websites in the search results, given that, in general, websites returned on the first Google search results page generate 92% of all traffic from an average search [16,33]. This drops by 95% for the second page and by 78% and 58% for subsequent pages [31]. Thus, we did not aim to identify all websites on statins but rather to emulate an authentic consumer search.

**Website Search Rank**

The effect of the association between search result ranking and accuracy, credibility, and readability was considered. Given that engagement is highest with the first 5 websites in search results, garnering 67% of all clicks from a search page [31,34], the websites were divided into 4 sets of 5 websites, each according to their ranking as per Google search result. Thus, websites ranked 1-5 were called quartile 1, websites 6-10 were called quartile 2, and so forth.

**Website Categorization**

In the interest of determining whether the nature of the authorship of the websites had a bearing on their accuracy, credibility, and readability, each of the 20 chosen websites was categorized into 3 types: commercial (defined as a website that generates revenue or cash and is not affiliated with the government), not-for-profit (a website that garners support for a cause rather than revenue, including government and charities), and media (a website that reports new findings or stories, with the primary purpose of the website being news reporting).

**Presence of HONcode Seal**

We assessed whether the quality of the websites was associated with the presence or absence of the HONcode seal.
## Table 1. Website accuracy checklist.

<table>
<thead>
<tr>
<th>Accuracy criteria</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mentions that cholesterol is a modifiable risk factor for cardiovascular disease [35]</td>
</tr>
<tr>
<td>2</td>
<td>Mentions that consultation with a doctor is essential before and while taking statins and when ceasing them [36]</td>
</tr>
<tr>
<td>3</td>
<td>Lists conditions for which statins are used [36]</td>
</tr>
<tr>
<td>4</td>
<td>Defines the target population for statin therapy [37]</td>
</tr>
<tr>
<td>5</td>
<td>Mentions the importance of adherence to statin therapy [36]</td>
</tr>
<tr>
<td>6</td>
<td>Addresses the subtleties of primary prevention [38]</td>
</tr>
<tr>
<td>7</td>
<td>Mentions that statins are about reducing complications of high cholesterol rather than achieving a specific (low-density lipoprotein) cholesterol [38]</td>
</tr>
<tr>
<td>8</td>
<td>Describes or at least lists the benefits of statin therapy [39]</td>
</tr>
<tr>
<td>9</td>
<td>Describes or at least lists the side effects or risks of statin therapy [39]</td>
</tr>
<tr>
<td>10</td>
<td>Describes how treatment affects the overall quality of life [39]</td>
</tr>
<tr>
<td>11</td>
<td>Mentions low to moderate dose statin therapy is recommended in primary prevention [37]</td>
</tr>
<tr>
<td>12</td>
<td>Specifically addresses rhabdomyolysis [36]</td>
</tr>
<tr>
<td>13</td>
<td>Describes the approximate financial burden to the patient [36]</td>
</tr>
<tr>
<td>14</td>
<td>Describes the duration of treatment before an effect is measurable [39]</td>
</tr>
<tr>
<td>15</td>
<td>Describes how statins work or at least what they do [39]</td>
</tr>
<tr>
<td>16</td>
<td>Describes what may happen without treatment [39]</td>
</tr>
<tr>
<td>17</td>
<td>Explores the possibility of using alternative therapies to statins [39]</td>
</tr>
<tr>
<td>18</td>
<td>Mentions that statins must not be used during pregnancy [37]</td>
</tr>
<tr>
<td>19</td>
<td>Describes drug interactions or at least lists them [39]</td>
</tr>
<tr>
<td>20</td>
<td>Mentions that statins do not replace a healthy lifestyle [36]</td>
</tr>
</tbody>
</table>

## Credibility

In developing the criteria to be included in the assessment of credibility (Table 2), DISCERN was chosen as a reference, as well as other studies that used DISCERN or another available tool for website assessment. However, as the 5-point Likert scale used in DISCERN can be subject to response style bias [40], a present (1) or absent (0) scale was adopted as it has been shown to improve the objectivity of data collection [41-43].
<table>
<thead>
<tr>
<th>Item</th>
<th>Criteria</th>
<th>Score criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Referencing or citations obtained from peer-reviewed journals</td>
<td>1 point if the articles for which the references are obtained are published in peer-reviewed journals [16,44]</td>
</tr>
<tr>
<td>2</td>
<td>Website updated within last 24 months</td>
<td>The latest update should be within the past 24 months [45]</td>
</tr>
<tr>
<td>3</td>
<td>Avoids anecdotal evidence for making claims</td>
<td>Does not use anecdotal evidence as a basis for claims; quoting a case study without using claims is acceptable [46]</td>
</tr>
<tr>
<td>4</td>
<td>Mailing address present</td>
<td>Physical contact address of the website clearly stated [45]</td>
</tr>
<tr>
<td>5</td>
<td>Contact information available</td>
<td>Contact information including name, position, telephone number, address, and email [47]</td>
</tr>
<tr>
<td>6</td>
<td>Sponsorship stated</td>
<td>Any sponsorship should be clearly stated</td>
</tr>
<tr>
<td>7</td>
<td>Organizational privacy policy stated</td>
<td>Organization privacy policy should be clearly stated [47]</td>
</tr>
<tr>
<td>8</td>
<td>Declaration of the author’s qualification</td>
<td>Author’s qualification should be health care related [16,44]</td>
</tr>
<tr>
<td>9</td>
<td>Paid access tab present</td>
<td>If paid access is available, the difference in the information obtained from paid vs unpaid access should be clearly stated [48]</td>
</tr>
<tr>
<td>10</td>
<td>Disclosure of funding or conflicts of interest</td>
<td>Conflicts of interest and funding disclosure should be clearly stated [44]</td>
</tr>
<tr>
<td>11</td>
<td>The presence of an HONcode seal or third-party certification</td>
<td>Presence of a HONcode seal or any other third-party certification [16]</td>
</tr>
<tr>
<td>12</td>
<td>Advertisement neutral</td>
<td>Advertisements should steer clear from the website information (eg, no pop-ups related to the website content) [16]</td>
</tr>
<tr>
<td>13</td>
<td>Disclaimer regarding web-based health information</td>
<td>A disclaimer should be clearly stated that web-based health information does not replace a practitioner’s advice [44]</td>
</tr>
</tbody>
</table>

*aHONcode: Health on the Net Code of Conduct.*

Each website was appraised according to this list. A score was allocated for each website’s front page, with internal links explored only if relevant. Data (credibility scores) were undertaken as independent assessments by 2 assessors (E Loh and DdP). The results were compared, and if discrepancies arose, discussions were held to clarify the score, with external input from advisors (HM, SL, and KS) obtained where appropriate.

**Readability**

For readability, various tools are available, including the Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, Simple Measure of Gobbledygook (SMOG), and Average Grade Level. We used SMOG as it is considered the gold standard for assessing the readability of health care material and has a high correlation with the other scoring systems [24,49]. Importantly, the outcome measure is easy to understand as, for example, an SMOG readability grade of 6 represents a text comprehensible to all individuals with sixth-grade reading skills and above [50-53]. This grade level was set as the basis of readability, given that the available literature sets this as the standard for “superior” readability. To use this tool, texts from the 20 selected websites were copied and saved as separate Microsoft Word (Microsoft Corp) and plain text documents for analysis, deleting text unrelated to the health information topic (eg, author information or disclaimers) to prevent this from confounding the scoring. A single web-based readability calculator [54] was used to generate the scores.

**Data Analysis**

The website category and ranking findings were compared by ANOVA, and differences between websites with and without the HONcode seal were analyzed with 2-tailed t tests. In addition, the relationship between credibility and readability with accuracy was assessed by Pearson correlation.

**Ethical Considerations**

As the research was not conducted on human subjects, no ethics review was required.

**Results**

**Selected Websites**

The top 20 websites returned by the search are listed in Multimedia Appendix 1 [55-74]. Of the 20 websites chosen from the search, 45% (n=9) were categorized as commercial, 45% (n=9) not-for-profit, and 10% (n=2) media (Table 3). There was an even distribution of commercial and not-for-profit websites across the 4 quartiles, with both media websites found in the fourth quartile. Eight of the websites bore the HONcode seal.
Table 3. Top 20 statin websites’ category, HONcode\textsuperscript{a} presence or absence, accuracy, credibility, and readability.

<table>
<thead>
<tr>
<th>Website rank</th>
<th>Category</th>
<th>HONcode Seal</th>
<th>Accuracy</th>
<th>Credibility</th>
<th>Readability (SMOG\textsuperscript{b} score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Commercial</td>
<td>Yes</td>
<td>20</td>
<td>6</td>
<td>10.7</td>
</tr>
<tr>
<td>2</td>
<td>Not-for-profit</td>
<td>Yes</td>
<td>9</td>
<td>11</td>
<td>11.3</td>
</tr>
<tr>
<td>3</td>
<td>Not-for-profit</td>
<td>No</td>
<td>23</td>
<td>9</td>
<td>12.5</td>
</tr>
<tr>
<td>4</td>
<td>Commercial</td>
<td>Yes</td>
<td>16</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>Commercial</td>
<td>No</td>
<td>10</td>
<td>4</td>
<td>7.7</td>
</tr>
<tr>
<td>6</td>
<td>Not-for-profit</td>
<td>No</td>
<td>15</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Commercial</td>
<td>No</td>
<td>14</td>
<td>10</td>
<td>9.6</td>
</tr>
<tr>
<td>8</td>
<td>Not-for-profit</td>
<td>No</td>
<td>13</td>
<td>5</td>
<td>10.8</td>
</tr>
<tr>
<td>9</td>
<td>Not-for-profit</td>
<td>No</td>
<td>16</td>
<td>11</td>
<td>10.2</td>
</tr>
<tr>
<td>10</td>
<td>Commercial</td>
<td>Yes</td>
<td>11</td>
<td>10</td>
<td>8.6</td>
</tr>
<tr>
<td>11</td>
<td>Commercial</td>
<td>Yes</td>
<td>16</td>
<td>9</td>
<td>9.4</td>
</tr>
<tr>
<td>12</td>
<td>Commercial</td>
<td>No</td>
<td>13</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>Not-for-profit</td>
<td>No</td>
<td>17</td>
<td>9</td>
<td>8.9</td>
</tr>
<tr>
<td>14</td>
<td>Not-for-profit</td>
<td>No</td>
<td>8</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>Not-for-profit</td>
<td>No</td>
<td>8</td>
<td>7</td>
<td>10.3</td>
</tr>
<tr>
<td>16</td>
<td>Media</td>
<td>No</td>
<td>8</td>
<td>4</td>
<td>7.8</td>
</tr>
<tr>
<td>17</td>
<td>Not-for-profit</td>
<td>Yes</td>
<td>16</td>
<td>11</td>
<td>6.7</td>
</tr>
<tr>
<td>18</td>
<td>Commercial</td>
<td>Yes</td>
<td>19</td>
<td>8</td>
<td>9.7</td>
</tr>
<tr>
<td>19</td>
<td>Commercial</td>
<td>Yes</td>
<td>21</td>
<td>7</td>
<td>9.2</td>
</tr>
<tr>
<td>20</td>
<td>Media</td>
<td>No</td>
<td>14</td>
<td>2</td>
<td>8.1</td>
</tr>
</tbody>
</table>

\textsuperscript{a}HONcode: Health on the Net Code of Conduct.

\textsuperscript{b}SMOG: Simple Measure of Gobbledygook.

Accuracy

The mean website accuracy score was 14.35 (SD 4.43). In terms of accuracy, the 3 highest-scoring websites were Wikipedia (score of 23), Drugs.com (score of 21), and Medicine.net (score of 20). These were the only websites to achieve a score of 20 or above. No website contradicted any checklist criterion. The top 3 performing checklist criteria were related to side effects and statin mechanism of action (criteria 9, 12, and 15), with a score of “present and complete” for each of these criteria achieved by 18, 14, and 12 websites, respectively. Although side effects were covered to some degree in all websites, criteria about drug safety (criteria 18 and 19) were complete in only 8 and 7 websites, respectively. Other poorly performing criteria reflected the lack of detail about primary prevention (criteria 11 and 6), with a score of “absent” assigned to 19 and 17 of the websites, respectively.

Credibility

None of the sampled websites met all credibility criteria for a perfect score of 13. The mean score overall was 7.45, with a range of 2-11. Importantly, 12 websites referenced peer-reviewed journal articles as a source of information, and 15 avoided anecdotal evidence for making claims. Media and some commercial websites reported personal opinions. While only 6 websites provided an organization’s contact details, the others provided an email address or feedback form for contact purposes. Sponsorship was explicitly stated in 12 websites, with reference to either government or private organizations. All websites declared their organization’s privacy policy, including websites with lower overall credibility scores. Only 8 websites declared author qualifications, which were primarily health related. None of the websites required paid access. Funding sources were fully disclosed in 11 websites, with the remaining 9 not reporting their source of funding or conflicts of interest. Twelve websites either had no advertisements or non–health care advertisements; the 8 websites that did not meet this criterion were commercial or media websites. Only 5 websites did not include a disclaimer that web-based health information does not replace a practitioner’s advice: all of these websites were commercial or media websites.

Readability

Overall, for SMOG readability, the median was 9.65 (IQR 8.825-10.85) and the average was 9.875 (SD 1.75), that is, above the ninth-grade level. None of the websites met the recommended grade level of 6; even the media websites required an eighth-grade level of comprehension.
Correlation Between Accuracy, Credibility, and Readability

No significant correlation was evident between the correlation between credibility and accuracy ($P=.23$) (Figure 1).

Figure 1. Relationship between website credibility (black triangles) and correlation readability (blue squares) with accuracy ($P=.23$).

Website Search Rank

Websites that featured prominently in search results were not necessarily the most accurate, with no significant difference between the quartiles ($P=.64$). Indeed, of the 20 websites reviewed, the second highest scoring website for accuracy was ranked in the 19th position on search results. Similarly, there was no difference in credibility ($P=.63$) or readability ($P=.06$) between the quartiles.

Website Categorization

Comparing commercial with not-for-profit websites, 2-tailed $t$ tests revealed there was no significant difference in terms of accuracy ($P=.275$), credibility ($P=.83$), or readability ($P=.452$). As there were only 2 media websites, a comparison with them was not made. Notably, they had the lowest scores for credibility, but both scored among the most readable.

Presence of HONcode Seal

Of the 20 websites, 8 were HONcode certified, with 6 of these categorized as commercial websites and 2 as not-for-profit. The mean accuracy scores for websites with and without the HONcode seal were 16 (SD 4.2) and 13.25 (SD 4.4), respectively, but this was not significantly different ($P=.18$). The presence or absence of the HONcode seal did not preclude a website from scoring at either end of the accuracy scale. Although the 8 websites with the HONcode seal scored higher in credibility (mean 8.5, SD 2) than websites that were not HONcode certified (mean 6.75, SD 2.7), this was not significantly different ($P=.139$). There was no significant difference in readability scores ($P=.92$) when comparing websites with HONcode seal status or lack thereof (mean 9.83, SD 1.9 and mean 9.91, SD 1.7, respectively).

Discussion

This study found that overall, the quality of websites with statin-related information tended to be poor. The website content was not sufficiently comprehensive, and the reading level was too difficult for the average reader to fully comprehend. The credibility of the websites varied, although overall websites bearing the HONcode seal had higher credibility than those without.

Here, we formally assessed the quality of websites addressing statins and their side effects. The finding that the quality of information is of variable caliber is consistent with studies investigating web-based health information on other topics [4,17,75]. Although the criteria used by Google’s ranking algorithm is confidential, Google’s guidelines state that it uses a series of algorithms that account for the words of the query, relevance and usability of web pages, the expertise of sources, ease of use on mobile device interfaces, as well as location and settings to determine the results displayed [76,77]. However, this study demonstrates that the most prominent websites in the Google search ranking are not necessarily of high quality.

The lack of correlation between accuracy and credibility or readability is a concern if patients are using the information to understand their condition and take action related to it. Patients with poor health literacy may use inaccurate and untrustworthy information in deciding whether to see a health professional following the onset of symptoms or whether to undertake tests and treatments that may be unnecessary or recommended by health professionals [12]. Vulnerable populations are at higher risk of having poor health literacy and experiencing CVD.
While most websites analyzed as part of this study scored low in accuracy, this tended to be attributed to a lack of completeness of information rather than a lack of factual information. While the checklist developed here may be stringent, it would be reasonable to expect that websites dedicated to statins would be comprehensive. The lack of comprehensiveness in the information provided on the websites could result in consumers overlooking important details unless they browse through multiple websites. Furthermore, visits to multiple websites may not generate clarity but confusion. This is due to the increased likelihood of encountering inconsistent information, particularly as websites have different agendas based on the website type. That said, the commercial websites scored, on average, just as well as the not-for-profit websites, indicating that they can be a valuable source of information for consumers. It also indicates that government and other not-for-profit websites will be required to at least match the accuracy of commercial websites if they are to remain relevant in Google’s search algorithm, as having information-rich content is a factor that contributes to higher search rankings [78].

When browsing the internet, one would expect government and other not-for-profit websites to provide credible information. However, some of these websites returned relatively low credibility scores and overall were not significantly more credible than commercial websites. Over half of the 20 websites analyzed provided evidence-based information and avoided anecdotal evidence, increasing their credibility rating [16,44]; however, it was concerning that 5 websites provided information based on anecdotal evidence. As expected, media websites received low credibility scores as news articles about statins contained personal views and anecdotes. Other indicators of credibility were lacking by a large proportion of websites, in particular author qualifications and details about sponsorship, funding, and conflicts of interest [16,44]. Furthermore, many of the commercial and media websites included advertisements, including health-related advertisements [16], and 5 of them did not include the disclaimer that web-based health information does not replace a practitioner’s advice [44]. Thus, even patients with a degree of health literacy would find it difficult to accurately appraise the credibility of many of these websites on statins.

Given that the general public is unlikely to be fully equipped to gauge the credibility of web-based health information presented [12], clinicians could advise that patients identify the presence of the HONcode seal as this merits some confidence in the information presented [16]. However, the code does not necessarily imply that websites are comprehensive. Additionally, website developer application for the HONcode seal is voluntary, so high-quality websites may not bear the HONcode seal. The finding that the readability of the websites with the HONcode seal was not at a suitable level means that such websites may not represent digestible patient health information. Furthermore, a practical issue is that the HONcode seal is at the bottom of the web page and is thus not necessarily evident at first glance.

Many patient demographic groups have been found to read at a level more than 3 years below their completed educational years [79]. Thus, the study results may not be indicative of the severity of the problem posed by websites with high readability scores in terms of the general public’s understanding of web-based information [79]. Those with limited literacy skills tend to have poorer health status due to a lack of knowledge and understanding of health care issues and a diminished ability to participate in shared decision-making in the clinical context [80]. They also tend to have poorer compliance with treatment recommendations and subsequent disease progression, as well as a higher risk for seeking emergency care and more frequent and longer inpatient admissions [27].

Additionally, other factors besides readability play into the way a text is received, including logical and sequential presentation of information. Additionally, alternative media such as images and graphs provide a well-documented “picture superiority” effect that boosts understanding of and engagement with a text [81], although some of these may also require interpretation by consumers.

A limitation of this study of these websites is that the internet is dynamic, with websites updated at any time. The search used in undertaking the study is constrained temporally in its noniterative nature, as well as its method, which used only the major search engine Google. While a metasearch capturing results from multiple search engines would provide a more comprehensive view of the information about statins on the internet, it is unlikely to represent the behavior of the public [82]. Additionally, only 8 of the websites in the study were updated in some way after completion of this study, and the information on some websites is dated as more than 10 years old.

Overall, this study has demonstrated that within the surfeit of information available on the internet regarding statin therapy, the quality of websites is of mixed caliber. The content of information is generally accurate but incomplete, while credibility is variable. Readability is generally of a level too difficult for the general public to comprehend. This suggests a need for guidance to website developers of health care websites in order to capitalize on the vast potential of the internet to equip patients with the empowerment of improved health information and health literacy. It also highlights that clinicians will need to be educated themselves about what is on the internet and what constitutes accuracy, credibility, and readability in order to impart this knowledge to their patients. During the COVID-19 pandemic, the methods through which patients seek information about their health have shifted toward increasingly internet-based means, making the quality of information on the internet of particular significance in the current climate and for the foreseeable future.
Acknowledgments

The findings were originally developed for the award of a Doctor of Medicine. Publication charges were funded by New South Wales (NSW) Health Pathology.

Authors' Contributions

E Ling attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. The project was planned by HM with input from all authors. Checklists were developed by E Loh and DdP with the assistance of SL and KS. Data were collected by E Ling, E Loh, and DdP. Data analysis was done by E Ling, E Loh, and DdP with assistance from HM. Data interpretation and manuscript review were done by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Top 20 websites analyzed in the study.

References


Original Paper

Online Visibility and Scientific Relevance of Strabismus Research: Bibliometric Analysis

Aleksander Stupnicki¹; Basil Suresh¹, iBSc (Hons), MBBS; Saurabh Jain¹,², MB BS

¹University College London Medical School, London, United Kingdom
²Royal Free London NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Aleksander Stupnicki
University College London Medical School
4 Huntley Street
London, WC1E 6DE
United Kingdom
Phone: 44 02031088235
Email: zchatup@ucl.ac.uk

Abstract

Background: Quality and accuracy of online scientific data are crucial, given that the internet and social media serve nowadays as primary sources of medical knowledge.

Objective: This study aims to analyze the relationship between scientific relevance and online visibility of strabismus research to answer the following questions: (1) Are the most popular strabismus papers scientifically relevant? (2) Are the most high-impact strabismus studies shared enough online?

Methods: The Altmetric Attention Score (AAS) was used as a proxy for online visibility, whereas citations and the journal’s impact factor (IF) served as a metric for scientific relevance. Using “strabismus” as a keyword, 100 papers with the highest AAS and 100 papers with the highest number of citations were identified. Statistical analyses, including the Spearman rank test, linear regression, and factor analysis, were performed to assess the relationship between AAS, citations, a journal’s IF, and mentions across 18 individual Web 2.0 platforms.

Results: A weak, positive, statistically significant correlation was observed between normalized AAS and normalized citations ($P<.001; r=0.27$) for papers with high visibility. Only Twitter mentions and Mendeley readers correlated significantly with normalized citations ($P=.02$ and $P<.001$, respectively) and IF ($P=.04$ and $P=.009$, respectively), with Twitter being the strongest significant predictor of citation numbers ($r=0.53$). For high-impact papers, no correlation was found between normalized citations and normalized AAS ($P=.12$) or the IF of the journal ($P=.55$).

Conclusions: While clinical relevance influences online attention, most high-impact research related to strabismus is not sufficiently shared on the web. Therefore, researchers should make a greater effort to share high-impact papers related to strabismus on online media platforms to improve accessibility and quality of evidence-based knowledge for patients.


KEYWORDS
strabismus research; squint; social media; scientific relevance; altmetrics; accuracy; medical knowledge; metric; bibliometric analysis; research; strabismus; online visibility; platform; evidence-based information; accessibility

Introduction

Patients, health care professionals, and researchers increasingly use social media and online platforms as a source of knowledge, health care news, and scientific research [1]. Despite the worldwide prevalence of strabismus remaining stable at around 2% [2], the public’s online interest in the topic has been rising, a trend reflected by the increasing popularity of queries related to the disease over the past 2 decades according to Google Trends. Due to this increasing reliance on online platforms, it is essential to ensure the quality and relevance of scientific data that are commonly accessed on the web, especially for lay members of the public who may lack the skills or time to assess that themselves.

To quantify the relevance of research within the field of medical science, the number of citations and impact factor (IF) of the
journal are used most frequently [3]. The dissemination of the same academic information through platforms used by the general public, on the other hand, can be most reliably quantified by the Altmetric Attention Score (AAS), a real-time weighted measure of mentions across all Web 2.0 social media platforms [4,5].

Bibliometric analyses using altmetrics and other scientometrics have been conducted previously in the field of ophthalmic research to evaluate publication trends [6], disruptiveness of papers [7], or research productivity [8]. To date, however, there have been no such analyses within the subspecialty of strabismus, despite the pervasiveness of the disease.

Therefore, we decided to analyze the relationship between the scientific relevance of strabismus research and its contributions to the online sphere, in order to answer the following questions: (1) Are the most popular strabismus papers scientifically relevant? (2) Are the most high-impact strabismus studies shared online enough?

Methods

In line with the best practice of literature searching [9], a thesaurus synonym search was performed to identify appropriate keywords for database search. As of January 2023, the thesaurus does not identify synonyms for “strabismus,” and the Cambridge Dictionary confirms it is the only medical term to describe the condition [10]. No morphological variation of the term has been identified, eliminating the need for the use of truncation in keyword searches.

Therefore, a list of research papers including the keyword “strabismus” was generated on January 27, 2023, with Altmetric Explorer with no other restrictions (search period: January 2011 to January 2023). The keyword search engine in Altmetric Explorer yields comprehensive results including outputs that match the keyword across publication title, author name, or journal title [11]. Hence, the pooled list was then filtered by a consultant ophthalmologist according to relevance to include 100 papers with the highest AAS (a total of 255 titles and abstracts were analyzed to compile 100 relevant publications). Additional preliminary searches using lay synonyms of strabismus, including “squint” and “cross-eye,” were performed but yielded no relevant or sufficiently high AAS results for inclusion, proving the keyword “strabismus” captures the bona fide core of publications in the field.

On the same day (January 27, 2023), for each of the papers, Web of Science (WoS) was used to add information on the number of citations, time since publication, and IF of the journal at the time of publication; other metrics traditionally used to assess the quality and relevance of scientific research [12]. Additional data on the source of AAS, including mentions across (1) news, (2) blogs, (3) Twitter, (4) peer review, (5) Facebook, (6) Wikipedia, (7) LinkedIn, (8) Weibo, (9) Google+, (10) Reddit, (11) Pinterest, (12) F100, (13) Q&A, (14) policy, (15) patent, (16) video, (17) syllabi, and (18) Mendeley, were pooled from Altmetric website and evaluated to characterize the field.

For systematic comparison, the same approach to searching was implemented to yield a list of papers with the highest number of citations: on January 27, 2023, WoS was used to generate a list of 100 papers including the keyword “strabismus” with the highest number of citations, excluding papers published before 2011, and the year Altmetric Explorer was founded and started tracking the AAS (search period: January 2011 to January 2023). No other filters were applied to the search. On the same day, the AAS for each of the papers was manually pooled from Altmetric website. Data on time since publication and the journal’s IF at the time of publication were extracted from WoS. To account for temporal differences [13], the values for AAS and citations for both groups have been then normalized per year since publication.

Kolmogorov-Smirnov test was used to verify that the distribution of the data does not follow a normal distribution, and Spearman rank correlation coefficient was used to test for correlation between all variables. Following correlational calculations, linear regression analysis and factor analysis were performed to explain patterns among correlated variables, both of which are statistical techniques commonly used in altmetric research [14,15]. SPSS (IBM Corp) was used for all statistical calculations. Statistical significance was defined as P<.05.

Ethical Considerations

No ethics board approval was required, as the study did not involve any human participants.

Results

Correlation Analysis

The normalized AASs of the 100 papers with the highest online visibility (median AAS 11, IQR 6-16) correlated significantly with normalized citations (P<.001) but demonstrated a weak strength of the relationship (r=0.27) for papers with AAS <150. To achieve this result, we excluded 3 outlier papers with significantly higher AAs (922, 413, and 169, respectively, compared to median 11, IQR 6-16; z score>3), which would otherwise skew the statistical analysis. Spearman rank test demonstrated no correlation between the normalized AAS and the IF of the journal (P=.15) or time (P=.37).

For the 100 papers with the highest number of citations (median 30, IQR 15-45), no statistically significant correlation was found between normalized citations and the normalized AAS (P=.12) or IF of the journal (P=.55), but as expected, they correlated significantly with time (P=.01).

Upon analysis of AAS sources, we found a weak, positive, statistically significant correlation between normalized citations and Twitter mentions (P=.02; r=0.27), normalized citations and Mendeley readers (P<.001; r=0.40), and normalized citations and policy mentions (P=.02; r=0.24) for the 100 papers with highest AAS. The same variables showed a weak, positive, statistically significant correlation with the IF of the journal at the time of publication: Twitter and IF (P=.04; r=0.25), Mendeley readers and IF (P=.009; r=0.32), and policy mentions and IF (P=.04; r=0.26).

Correlations between the number of mentions in the news, on blogs, in peer-reviews, on Facebook, Wikipedia, LinkedIn, Reddit, Google+, Weibo, Pinterest, syllabi, or video and...
Multiivariate Analysis
To better understand variance among the correlated variables, a linear regression model was run with normalized citations as the dependent variable and Twitter mentions, Mendeley readers, and policy mentions as covariates. ANOVA test showed significant variance within the sample, confirming the suitability of the test ($P<.001$). We obtained an $R^2$ value of 0.31, indicating that 31% of the variance within citations can be explained cumulatively by the 3 AAS sources. Only Twitter and policy mentions, however, were significant predictors ($P<.001$ and $P=.004$ respectively), with Twitter mentions being the most important predictor as indicated by the highest standardized coefficient ($r=0.53$).

Factor Analysis
Factor analysis was performed on metrics with adequate data for the papers with the top 100 AAS normalized citations, normalized AAS, IF, time since publication (months), Twitter mentions, Mendeley readers, and policy documents. Bartlett test of sphericity indicated an approximate chi-square value of $\chi^2_{93}=124.4$ ($P<.001$) and a Kaiser-Meyer-Olkin adequacy value of 0.607, together indicating the suitability of the data set for factor analysis. Three factors were identified across these variables: factor 1 between Twitter, Mendeley, AAS, and citations; factor 2 between Mendeley, policy, and citations; and factor 3 between IF and time.

Discussion
Principal Findings
The significant, yet, weak correlation of AAS with citations for papers with the highest online visibility shows that the clinical relevance of strabismus-related publications (as measured by citations) can contribute to increased online popularity but is not the sole determining factor. Furthermore, the lack of correlation between AAS and IF demonstrates that the relative importance of a journal in the field (and consequently the paper) does not determine its online popularity, which raises questions about the quality of strabismus research receiving the most online attention.

Through correlational and multivariate analysis of mentions across individual Web 2.0 platforms, we have demonstrated Twitter mentions to be a significant and strong predictor of citations for the most popular strabismus papers. Although prior studies demonstrated a more significant, causative impact of tweets on citation numbers [16], the relationship can differ between fields [17] and seems to be statistically significant for strabismus research, albeit moderate compared with other research domains. Overall, our findings imply that dissemination of strabismus research through Twitter can have an impact on scholarly visibility and subsequently citation rates.

Furthermore, we demonstrated a lack of statistically significant correlations between traditional scientometrics (citations and IF) and mentions across other social or media platforms, which has been also observed in other fields of research [17]. This reveals existing gaps that require more references to research papers on strabismus, including social media platforms like Facebook or LinkedIn, as well as critical, knowledge-oriented pages such as Wikipedia.

Furthermore, in terms of factor analysis, factor 1 linking AAS, citations, Mendeley, and Twitter likely suggests that for strabismus research, there is a degree of overlap regarding the user bases or networks between Mendeley and Twitter, despite the former being considered a platform largely used by academic professionals as opposed to the more widely public microblogging service [18]. It may also indicate that highly cited papers are receiving engagement and being discussed in both academic and general public networks indicating that such papers may have a wider social impact. Factor 2 linking Mendeley, policy documents, and citations may suggest that for highly cited papers, there is increased interest and readership on Mendeley—a proportion of which may be faculty and departmental figures. This in turn may lead to policy mentions for impactful research papers. Therefore, this suggests that papers with high citation counts and academic impact may be influencing policies and organizational standards [19]. Factor 3 linking IF and time could be due to the overall increase and growth of the cited strabismus literature over time; however, this is less relevant to our research question.

For the 100 papers with the highest number of citations, the lack of correlation between citations or IF and the AAS suggests that clinical relevance or perceived prestige related to the publishing journal does not affect the online visibility of strabismus papers. Researchers publishing in the strabismus realm should, therefore, make a greater effort to share their high-impact papers on social media. In turn, this could increase the visibility and accessibility of their research, especially for the lay public who rarely browses journals for medical knowledge, enhance collaboration, and further enhance the overall impact of their research.

Strengths and Limitations
AAS itself is a useful tool for authors to get quick, up-to-date insight into the performance of their papers on the web. It is crucial, however, to bear in mind the inherent limitations of AAS due to the fact that it is ultimately only a metric of “mentions” or “posts” and is not an indicator of research quality or legitimacy [20]. In isolation, it may be deemed unreliable, as “viral” papers that do not exhibit robust research methodology or present sensible conclusions may still acquire a high AAS. Furthermore, the AAS does not account for following or website traffic, therefore providing no information on the actual number of viewers. As a result, a frequently mentioned paper can effectively have low visibility and reception, despite a high AAS. Although Twitter mentions and Mendeley readership seem to have some impact on citation numbers of the most popular strabismus papers, they only account for a small proportion of the variance within citation data (31%). Thus, using alternative metrics, like tweets, as predictors of scientific contribution and success does not constitute a comprehensive and precise appraisal method, as demonstrated before across several other fields [16,17].
Furthermore, citations take a longer time to accumulate, whereas AAS is updated almost in real time, so even when a new paper is published that can have high publicity in social media and among the scientific community, the citation numbers will lag months or sometimes years behind. Due to this phenomenon, there may be weaker or no association between citations and AAS for the latest research papers, which could skew the results of our 12-year view. Further statistical testing would be necessary to confirm that.

Additionally, in the case of strabismus research, both the AAS and citation numbers are characteristically low, which raises questions about the reliability of the data set, as it is analogical to having a small sample size. Various scientifically irrelevant factors can cause a high AAS, such as the topic of the paper, sensationalism, how easily it is understood by the general public, or the number of intersections of the topic with other branches of medicine. A good example of the effect of those confounding factors is a publication included in our data set entitled “Evidence That Leonardo da Vinci Had Strabismus,” which had the highest AAS of 922 (over 83 times the median score) but only 6 citations. This demonstrates that especially for papers with high AAS scores, the virality of the topic can have a higher impact on the AAS than its scientific significance.

Conclusions and Future Directions

We have demonstrated that the clinical relevance of strabismus research contributes to the amount of online attention it receives. However, the most high-impact strabismus research is not sufficiently shared across online platforms. Therefore, we recommend that researchers make a greater effort to share high-impact studies on social media platforms to improve the quality of evidence-based information about strabismus and improve the accessibility of this knowledge. To maximize the societal impact of research, it is important to interact with both academic and general audiences, as shown by the overlap between Mendeley and Twitter engagement of strabismus publications.

Furthermore, we revealed Twitter mentions to be the strongest predictor of citation numbers for strabismus papers, highlighting the potential impact of social media on scholarly visibility. Our findings also highlight the need for engagement of strabismus researchers across a broader range of platforms, including Facebook, LinkedIn, or Wikipedia. However, due to its inherent biases and limitations, the AAS itself or mentions across specific platforms should only complement traditional metrics, such as IF and citations, to provide a broader picture of the publicity of the paper but should not act as a stand-alone metrics for assessing the quality and relevance of strabismus papers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary table of Spearman rank test for normalized citations, impact factor, and all 18 Web 2.0 platforms analyzed. [XLSX File (Microsoft Excel File), 11 KB - ijr_v13i1e50698_app1.xlsx ]

References


15. Taherdoost H, Sahibuddin S, Jalaliyoon N. Exploratory factor analysis; concepts and theory. SSRN J 2022 [FREE Full text]


Abbreviations

AAS: Altmetric Attention Score
IF: impact factor
WoS: Web of Science

©Aleksander Stupnicki, Basil Suresh, Saurabh Jain. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 12.06.2024. 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.
Census-Dependent Mortality of Ventilated Patients With COVID-19 in Israel: Noninterventional Observational Cohort Study

Joseph Mendlovic, MHA, MD; Francis B Mimouni, MD

1Shaare Zedek Medical Center, Affiliated with the Hadassah-Hebrew University School of Medicine, Jerusalem, Israel
2Associate Director General Office, Israel Ministry of Health, Jerusalem, Israel
3Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
4Leumit Health Service, Tel Aviv, Israel

Corresponding Author:
Joseph Mendlovic, MHA, MD
Shaare Zedek Medical Center
Affiliated with the Hadassah-Hebrew University School of Medicine
PO Box 12000
Shmuel (Hans) Beyth St 12
Jerusalem, 91120
Israel
Phone: 972 6777111 ext 1
Email: sefimen@gmail.com

Abstract

Background: The COVID-19 pandemic led to several surges in the mass hospitalization rate. Extreme increases in hospital admissions without adequate medical resources may increase mortality. No study has addressed the impact of daily census of ventilated patients on mortality in the context of the pandemic in a nationwide setting.

Objective: This study aimed to determine whether daily census of ventilated patients affected COVID-19 mortality rates nationwide in Israel.

Methods: We conducted a cohort study using nationwide, public-domain, population-based COVID-19 data of hospitalized patients from an Israeli database from March 11, 2020, until February 11, 2021. We included all COVID-19 hospital admissions, classified as mild to severe per the Centers for Diseases Control and Prevention classification irrespective of whether they were mechanically ventilated. Outcome measures were daily death rates and death rates expressed as a percentage of ventilated patients.

Results: During the study period (338 days from March 11, 2020, to February 11, 2021), 715,743 patients contracted and were clinically confirmed as having COVID-19. Among them, 5577 (0.78%) patients died. In total, 3398 patients were ventilated because of severe COVID-19. Daily mortality correlated with daily census of ventilated patients ($R^2=0.828$, $P<.001$). The daily percent mortality of ventilated patients also correlated with the daily census of ventilated patients ($R^2=0.365$, $P<.001$)—backward multiple regression analysis demonstrated that this positive correlation was still highly significant even when correcting for the average age or gender of ventilated patients ($R^2=0.4328$, $P<.001$) or for the surge in their number. Overall, 40% of the variation in mortality was explained by variations in the daily census of ventilated patients. ANOVA revealed that at less than 50 ventilated patients per day, the daily mortality of ventilated patients was slightly above 5%, and it nearly doubled (10%) with 50-149 patients; moreover, in all categories of ≥200 patients ventilated per day, it more than tripled at ≥15% ($P<.001$).

Conclusions: Daily mortality rates per ventilated patient increased with an increase in the number of ventilated patients, suggesting the saturation of medical resources. Policy makers should be aware that expanding medical services without adequate resources may increase mortality. Governments should perform similar analyses to provide indicators of system saturation, although further validation of these results might be needed to use this indicator to drive public policy.


KEYWORDS
COVID-19; mortality; ventilation; intensive care; pandemic; contagious; disease; mortality; database; data; patient; mortality; medical; resources; validation; public policy; policy; pandemic; health policy; global health policy
Introduction

SARS-CoV-2, the causative agent of COVID-19, was first identified on November 17, 2019, and was declared a pandemic by the World Health Organization on March 11, 2020 [1]. The virus is extremely contagious and the emergence of multiple mutated strains increased its contagiousness [2]. The disease’s severity, as observed in a small percentage of patients, warrants complex intensive care facilities [2]. At the onset of the pandemic, medical systems worldwide have been tremendously challenged by COVID-19, leading to major disruptions in routine hospital services, leading to chaos and exhausting reserve medical supplies [3-6]. Excess mortality beyond expected rates has been observed in many countries, including Israel [7]. Possibly, many potentially curable patients might have died because medical services were overwhelmed [6-8]. However, the extent of the impact of the saturation of medical services on a country’s COVID-19 mortality has not been systematically studied, and no study has addressed the impact of census on mortality in the context of the pandemics in a nationwide setting. The existence of such a relationship may be suggested by the strong correlation reported between the number of hospital beds per population size and COVID-19–specific mortality both in the United States [9] and worldwide [10]. In Israel, the number of intensive care units (ICU) beds is much lower than that in other high-income countries such as Germany or the United States [11,12].

We analyzed the national database of the Israeli Ministry of Health (MOH), systematically collected and reported since March 11, 2020, until February 11, 2021, that is, prior to mass vaccination conducted in Israel.

We hypothesize that daily COVID-19 mortality rates would be directly related to the daily national census of ventilated patients, increasing each time with an increase in the number of ventilated patients. Alternatively, it was also possible that because of routine experience in treating ventilated patients in regular wards, there would be no discernable effect of workload on mortality.

Methods

Study Design

In this retrospective cohort study, data were extracted from a nationwide, official, open-access COVID-19 database of Israel [13] and the MOH [14], specifically constructed for the purpose of reporting and research.

Setting

We used a national, public repository database, which was curated and made available on a government website to the general public for free use.

We collected data from all 31 general public hospitals in Israel, whereby clinical data were reported to the MOH 3 times daily. Data were uploaded automatically through a dedicated interface. After processing and quality control testing, it became possible to obtain a nationwide perspective of hospitalized patients. Simultaneously, with automatic transfer of information, manual reports were also transmitted from hospitals for the purpose of quality control, as described below. Data for a particular data were finalized, and updated data were uploaded into the system by midnight of any particular day. These databases were updated every weekday, 3 times daily. The data included the daily cumulative number of patients having tested positive for COVID-19, new laboratory-proven cases, vaccinated people (newly and cumulative), hospitalized patients (with mild, moderate, or severe COVID-19), mechanically ventilated patients with their average age and gender, and COVID-19–related deaths.

In addition to hospital-generated data, laboratory data of all new patients were regularly uploaded to the database through specific interfaces between testing laboratories and MOH computers. In Israel, SARS-CoV-2 positivity was confirmed via polymerase chain reaction–based swab tests using samples obtained from both the throat and one nostril in each patient. Testing was conducted at designated testing sites, such as task-specific health maintenance organizations’ clinics converted to testing sites, task-specific testing tents, mobile vans, and Home Front Command run testing compounds, in association with Magen David Adom (the Israeli “Red Cross”) and health maintenance organizations. In addition, patients in isolation or those who were unable to reach the testing site for medical reasons were tested at home. These laboratories reached a peak of >120,000 tests per day, while the Israeli population is approximately 9,300,000 [13].

To fully understand the settings of this study, one must be aware of how the pandemic was handled in Israel. The Israeli national health system is highly centralized and is under strict MOH regulation. During the pandemic, and in view of the chaos publicized in many countries whose medical systems were overwhelmed, the MOH took the following measures (among others): (1) throughout the whole pandemic, it facilitated the redirection of critically ill patients among medical centers to distribute the patient burden evenly; (2) all elective surgeries were canceled to focus hospital activities on patients with COVID-19—the latter requirement was strictly followed during the first peak of the pandemic, and was thereafter canceled to prevent potential harm; and (3) during surges in the number of hospitalized patients, care of critical patients (both ventilated and nonventilated) was provided by additional physicians and nurses who are not related to anesthesiology or ICU departments but have previous experience of handling ventilated patients in non-ICU wards; indeed, there are not enough ICU beds in Israel to take care of all ventilated patients, and many of them are cared for in non-ICU wards [15]. Owing to the scarcity of experienced physicians and nurses (some of whom were in quarantine due to unprotected, inadvertent exposure to SARS-CoV-2) during peaks of the COVID-19 pandemic, additional medical personnel were assigned to coronavirus wards after receiving a “crash” course on ventilator management. In Israel, ICU beds are usually occupied at a rate of approximately 100%, and the overflow is taken care of on regular wards. During the pandemic, most critically ill patients with COVID-19 ended up being cared for in satellite units (designated coronavirus wards) because of 100% occupancy rates of formal ICUs, and to facilitate the isolation of these patients. Importantly, in Israel, there are no hospitals such as the so-called
“community hospitals” (level 1 hospitals) present in the United States, which do not contain ICUs; thus, all hospitals contributing to this database are regional, tertiary care centers.

The period of data collection in this study was from March 11, 2020, when the database was implemented, until February 11, 2021. All hospitalized patients were followed up from exposure and until death or discharge from the hospital (hereinafter the “follow-up period”). As part of quality control processes, a comparison was made between automatic and manual reports. Gaps were checked manually and corrected as needed. Additional checks were carried out for deceased patients, and hospital reports and death certificates were additionally compared to verify that the main cause of death was COVID-19–related.

**Participants**

The participants in this study were all the patients who tested positive for COVID-19 at some point during the study period. They were selected in accordance with a laboratory-confirmed diagnosis as described above. The patients were classified by the severity of symptoms: mild, moderate, and severe. Initially, each hospital report used its own criteria for levels of severity. From July 12, 2020 (day 123 of the study period since the implementation of the database on March 11, 2020), the MOH issued specific definitions established by the National Institutes of Health [16,17]: ventilated patients were those requiring invasive mechanical ventilation (with an endotracheal or tracheostomy tube). Mild disease was defined as laboratory confirmation in combination with mild symptoms (fever, cough, weakness, and loss of taste and smell). Moderate disease was defined as laboratory confirmation together with pneumonia (clinical or through imaging). Severe disease was defined as laboratory confirmation with one or more of the following: a respiratory rate of >30 breaths per minute, oxygen requirement of >30%; oxygen saturation in arterial blood of ≤93% in ambient air, and a ratio of arterial oxygen pressure and oxygen requirement of <300.

**Follow-Up Methods**

The number of deaths was updated daily. The overwhelming majority of patients who died were invasively ventilated prior to dying, and several of them were even placed on mechanical ventilation during transport to the hospital. We cannot, however, rule out that some patients died prior to arriving at the hospital; nonetheless, they were customarily ventilated by paramedics en route to the hospital unless they belong to the very small group of “do not resuscitate” patients because they have an incurable disease that would potentially lead to their death within less than 6 months (according to the Israeli “Dying Patient Law”).

**Assessments**

The main outcomes of interest were whether the patient died or not, was ventilated or not, their length of stay (LOS) in the hospital, and the duration for which they received mechanical ventilation. Potential confounders, predictors, and effect modifiers were the only variables that were prospectively collected from this database, namely gender, patients’ age, and the daily census of ventilated patients in Israel.

**Data Source**

The data source is the Israeli MOH’s National COVID-19 Database, systematically collected and reported since March 11, 2020, until February 11, 2021, that is, prior to mass vaccination that occurred in Israel.

**Ethical Considerations**

The website we used is a public repository, available free of charge to the public. It is completely anonymized and deidentified; therefore, it was appropriate to not apply for approval from an ethics review board. Under the federal regulations for human subjects research (45 CFR Part 46), research involving publicly available data sets would not require review by an institutional review board—no application is required—as long as the data are obtained from sources that are publicly available and are deidentified and uncoded as in this study [18]. Nevertheless, since the data were collected in Israel, the National Committee for Human Medical Research of the Israeli MOH provided its full approval for the study and waived the requirement for obtaining informed consent. The Israeli National Committee for Human Medical Research deemed this study exempt from ethical approval since this study involves public data processing for the purposes of policy making and reflecting on the national system’s dealings with the epidemic; therefore, it does not require to adhere to the tenets of the declaration of Helsinki.

**Statistical Analysis**

The Minitab Statistical Package (version 16; Minitab, LLC) was used for analyses. Data were tested for normality and expressed as mean (SD) or median (IQR) values as requested. Stepwise backward multiple regression analysis was carried out to determine the correlation between daily percent mortality per group of ventilated patients (dependent variable) and the daily census of ventilated patients (independent variable), while taking into account potential confounding variables that may affect daily percent mortality, such as mean patient age, gender, and COVID-19 surge (surges 1, 2, or 3). This method places all the independent variables at once in the equation, and eliminates those found to be insignificant sequentially step by step, repeating the operation each time until only the significant variables remain in the final equation. This method allows for results that are not influenced by the order of introduction of the independent variables. Daily percent mortality per group of ventilated patients was calculated as the daily mortality rate in Israel divided by the census of ventilated patients on the same day. Independent variables entered in the regression equation were only those found to influence (in univariate analysis) the daily percent mortality of ventilated patients at an α value of <.10.

Analysis of means using ANOVA was carried out to determine differences in the mortality rates of ventilated patients by group based on their number in the daily census. We arbitrarily analyzed daily census according to 8 groups of increasing census size: group 1 containing 1–49 ventilated patients, group 2 containing 50–99 ventilated patients, group 3 containing 100–149 ventilated patients, group 4 containing 150–199 ventilated patients, group 5 containing 200–249 ventilated patients, group
6 containing 250-299 ventilated patients, group 7 containing 300-349 ventilated patients, and group 8 containing 350-399 ventilated patients.

No sample size calculation was carried out as we used all the available numbers nationwide.

**Results**

During the study period (338 days from March 11, 2020, to February 11, 2021) 715,743 patients contracted and had a laboratory-confirmed diagnosis of COVID-19. They constitute the study population. Among them, 5577 (0.78%) patients died.

In total, 3398 patients were ventilated because of severe COVID-19. We retrieved complete data (including outcomes, gender, age, and LOS in hospital) for 3373 of them (as of this writing, there was no determined outcome for 25 patients who are still ventilated). These data are presented in Table 1. Briefly, patients who died were older, and there were many more men ventilated than women, but mortality among ventilated men and women was similar (approximately one-third of patients). LOS in hospital was much shorter among patients who died than among those who survived. The LOS to ventilation (from hospital admission to the end of hospital stay) was shorter for patients who died, and the LOS from ventilation to outcome (discharge or death) was much longer for survivors.

**Table 1.** Demographic data for ventilated patients (N=3373).

<table>
<thead>
<tr>
<th>Gender (male:female), n:n (% males)</th>
<th>Length of stay to outcome (days), mean (SD; median; IQR)</th>
<th>Length of stay to ventilation (days), mean (SD; median; IQR)</th>
<th>Length of stay from ventilation to outcome (days), mean (SD; median; IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alive (n=1128; 33.44%)</td>
<td>35.7 (29.4; 27; 15-47)</td>
<td>4.9 (5.9; 3; 1.6)</td>
<td>30.8 (28.4; 22; 11-42)</td>
</tr>
<tr>
<td>Dead (n=2245; 66.56%)</td>
<td>18.7 (16.5; 15; 8-25)</td>
<td>6.0 (6.6; 4; 1-8)</td>
<td>12.8 (15.1; 8; 3-17)</td>
</tr>
</tbody>
</table>

\(a P<.001.\)

\(b P=.18.\)

**Figure 1** depicts the daily number of patients hospitalized every day of the study for severe disease and the number of ventilated patients, showing 3 surges of frequencies. **Figure 2** depicts the daily mortality per day of the study, which also shows 3 surges of daily death rates, parallel to the surges in the number of patients with severe disease and ventilated patients. **Figure 3** shows the daily percent mortality of ventilated patients over time, which also follows a similar pattern to the 3 peaks.

Daily mortality correlated with the daily census of ventilated patients (\(R^2=0.828, P<.001; \) Figure 4). The daily percent mortality of ventilated patients also correlated with the daily census of ventilated patients (\(R^2=0.365, P<.001; \) Figure 5). Backward multiple regression analysis revealed that the latter positive correlation was still highly significant even when correcting for the average age or gender of ventilated patients (\(R^2=0.4328, P<.001\)) or for the wave number (\(P>.05\)). Both average age and gender remained significant (\(P<.001\)) in the final analysis (older age and a higher percentage of females contributing to higher mortality; Table 2). The \(R^2\) value of the correlation equation predicting the percentage mortality per ventilated patient was at best 0.4, implying that 40% of the variation in mortality can be explained by variations in the daily census of ventilated patients. Obviously, the rest or the variability (60%) was explained by other factors that were probably patient-dependent (such as BMI, diabetes, or other chronic diseases), which were not retrievable from this database.

**Figure 1.** Daily rates of patients with severe COVID-19 (black dots) and those ventilated (red dots) over time.
Figure 2. Daily mortality over time.

Figure 3. Daily percent mortality of ventilated patients over time.

Figure 4. Daily mortality rates versus the daily census of ventilated patients.
Table 2. Multiple regression analysis results showing the relative contribution of the daily census of ventilated patients, gender, and age (independent variables) on the daily percent mortality of ventilated patients (dependent variable).

<table>
<thead>
<tr>
<th></th>
<th>Daily census of ventilated patients</th>
<th>Gender</th>
<th>Age</th>
<th>Model summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial $R^2$</td>
<td>0.3574</td>
<td>0.0275</td>
<td>0.0005</td>
<td>0.3915</td>
</tr>
<tr>
<td>Effect size (adjusted sum of squares)</td>
<td>3772.7</td>
<td>290.7</td>
<td>68.6</td>
<td>5012.5</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.09</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

All variables here refer to the dependent variable, that is, daily percent mortality of ventilated patients.

ANOVA (Figure 6) revealed that with <50 ventilated patients per day, the daily mortality of ventilated patients was slightly above 5%; with 50-149 patients, it nearly doubled (10%); and in all 3 categories of 200 and more patients, it more than tripled at ≥15%.

Figure 6. Analysis of the mean percent mortality of ventilated patients by ventilated patients’ groups (group 1: 1-49; group 2: 50-99; group 3: 100-149; group 4: 150-199; group 5: 200-249; group 5: 250-299; group 6: 300-349; group 7: 350-399).

We investigated whether an increase in the number of ventilated patients during a specific surge may have been related to an accumulation of mechanically ventilated patients from previous surges. In fact, LOS or length of ventilation did not differ among patients who were admitted during surges 1, 2, or 3 (ANOVA $P$>.05). Moreover, among 11 of 164 patients admitted to hospital during surge 1 died during surge 2, and none died during surge 3; 65 of the 1088 patients admitted to hospital during surge 2
died during surge 3 (together with 993 patients admitted to hospital during surge 3)—among them, the vast majority (n=63) were admitted within the last 20 days of surge 2.

Discussion

Principal Findings

This is the first nationwide study demonstrating increasing mortality with increasing demand for health care resources during the COVID-19 pandemic. Indeed, daily mortality rates expressed as a percentage of ventilated patients correlated with the daily census of ventilated patients. We do not believe that this increase in percent mortality rates during surges was due to the accumulation of mechanically ventilated patients who stayed alive until but died during the following surge because the LOS or length of ventilation until death did not differ among patients who were admitted during surges 1, 2, or 3, and also because such “leftovers” were much less than 10% of a given surge.

In China, higher mortality was recorded in Wuhan than in other Chinese provinces, which is related to the rapid escalation in the number of infected patients with insufficient access to health care resources [19]. In the United States, a retrospective cohort study of a 26-hospital–integrated delivery system showed an association between a greater percentage of COVID-19–related admissions of hospital capacity and a lower survival rate [20]. There are multiple examples in the news media describing “accidental deaths” due to improper care of patients with COVID-19 [21]. To limit the dimensions of this catastrophe, many countries had imposed lockdowns and quarantines, leading to consecutive peaks of the pandemic [21].

Israel is generally considered to be technologically sophisticated [22], but its medical system is greatly stretched [23]. In a 2020 report from the 37 nations belonging to the OECD (Organisation for Economic Co-operation and Development), Israel features among the bottom 4 nations in terms of acute hospital care beds (2.2 per 1000 population), which is well below countries such as Japan (7.8 per 1000 population) [23]. Israel also features among the 11 countries with the lowest numbers of both physicians and nurses per 1000 population [23]. The Israeli system considers itself efficient, reporting daily occupancy rates among the highest worldwide, at 93.3% on average, second only to Ireland among the 37 OECD countries [23]. In Israel, the number of ICU beds is 4.6 per 100,000 population and 2.2 ICU beds per 100 hospital beds (reported in 2007) [11], which is much lower than that in other high-income countries such as Germany (24.6 hospital beds per 100,000 population and 4.1 ICU beds per 100 hospital beds), or the United States (20 hospital beds / 100,000 and 9 ICU beds per 100 hospital beds) [12]. In Israel, every winter, many patients are ventilated in internal medicine wards because of the unavailability of ICU beds [15]. However, in non–COVID-19 times, the mortality of ventilated patients in non-ICU wards was higher than that of formal ICU beds, proving that the system might be “cheaper” but not necessarily better [15]. There is a strong correlation between the number of hospital beds per population size and COVID-19–specific mortality both in the United States [9] and globally [10]. This does not imply that more is necessarily better. Indeed, case-mix and ICU organization are important to consider. Strategies for how to use ICU beds, the proportions of mechanically ventilated versus nonventilated patients, the availability of intermediary care, and other factors are important to consider. For instance, if a country has a policy of admitting nonventilated patients to high- or medium-care departments instead of ICUs or transferring extubated patients to intermediary care units immediately, such a country will need fewer ICU beds than those that do the opposite.

Daily mortality expressed as a percentage of ventilated patients correlated with the daily census of ventilated patients. This was true even after taking into account the average age of ventilated patients and their gender or surge number. We suggest that the medical system reached saturation, that is, its inability to adequately handle complex ventilated patients. The reasons for this saturation are multiple. We cannot currently determine whether or not there was a shortage of specific drugs, mechanical ventilators, oxygen, or other supplies, or a shortage of personnel, higher patient-nurse ratios, lack of ICU beds, etc, but these have been reported anecdotally in the general nonmedical literature [24]. In fact, it is highly possible that whenever the census of ventilated patients increased, some patients died because they could not receive proper care and not because their disease was incurable. These data should be considered within the context of the quasi-heroic behavior of exhausted, overworked teams of caretakers that performed their duties throughout the pandemic under perilous circumstances. Furthermore, it is highly possible that parallel increases in the mortality of mechanically ventilated patients without COVID-19 occurred during the same periods. Unfortunately, no similar database for was maintained for patients without COVID-19, and we were not able to verify if this occurred.

In this study, mortality was observed among approximately two-third of patients—a number difficult to compare to that of other countries since reported death rates of ventilated patients with COVID-19 is also dependent upon case-mix, varying from 48% among younger patients (younger than 40 years) to 84% among older ones (> 80 years of age) [25]. It has recently been shown that for instance, the physical manifestations of frailty and comorbidity, particularly a history of cognitive impairment and falls, may be useful in identifying patients with COVID-19 who need additional support during hospitalization and may be at a higher mortality risk [26]. This may also be dependent on noninvasive and invasive ventilation strategies. A recent meta-analysis by Lim et al [26] estimated the mortality rate among such patients to reach 45% on average, which is lower than that reported in this study. In our study, the effect of the census of ventilated patients census on daily mortality was nearly identical during the 3 different surges, which suggests that the SARS-CoV-2 variants, likely to differ among the various waves, affected mortality in a similar manner.

Worldwide, there have been reports of “spontaneous” reductions in critical care admissions, such as those related to stroke and cerebral emergencies [27], accidents, emergency surgery, and acute coronary events [28-30], which may have reduced the ICU burden. We suspect that it may have existed in Israel as well in particular during the first surge, but it did not reduce the ICU burden to a point that prevented an increase in mortality...
rates reported here. A recent study reported that the management of critically ill patients with COVID-19 in the United Kingdom was far from ideal in numerous cases, with systematic errors in the measurement of height and derived ideal body weight and delayed applications or nonimplementation of evidence-based interventions for acute respiratory distress syndrome (in particular, prone positioning) [31].

Our findings may not be exclusive to the COVID-19 pandemic. For instance, Israel has not yet achieved peace with all its neighbors, and a large-scale increase in hostilities may have led to a number of casualties that the Israeli medical system might not be able to handle. Israel is also located on The Great Syria-African Rift and is at risk of major and potentially deadly earthquakes to occur.

The major strengths of this study are the use of a national large database, a long study period for most part of the pandemic, and the measures undertaken for quality control. A limitation of this study relates to changes in some definitions in the middle of data collection. These changes are unlikely to influence our main findings, in that mortality rates (regardless of severity staging) correlated with the workload of ventilating patients. Another limitation of this study is that the data set did not evaluate workload in individual hospitals and individual ICUs (occupancy rates, staffing patterns, and hours of nursing care per patient per day) relative to resource availability, which somewhat limits drawing causal inferences. It is unclear whether our findings are universal or change when examined in accordance with a hospital’s geographic location or size; however, of note, during each surge, the Israeli MOH frequently intervened and helped individual hospitals to move patients between hospitals to prevent maldistribution. We also provide no data on ventilated patients in non–COVID-19 beds, whose survival might also have been affected by the shift of many medical resources from other departments to care for patients with COVID-19; hence, we were unable to provide data on noninvasive ventilation or the use of vasoactive drugs or even extracorporeal membrane oxygenation that was administered to some patients. This information would have potentially helped stratify the complexity of health care requirements but was not available in this database. Finally, outcomes noted were either discharge from hospital or death, while some patients may have died after discharge. We do not believe that withholding strategies due to lack of bed availability may have influenced survival since in Israel, since patients requiring ventilation are not dependent upon bed availability, and are ventilated in non-ICU settings. Additionally, the law in Israel does not allow extubating of a dying patient.

**Recommendations to the Health Care Leadership**

Recognizing the fact that the ability to provide adequate intensive care and respiratory care services is a critical and unique national resource during pandemics and other emergencies. The use of percentage mortality or the census of ventilated patients as a potential key tool for monitoring the hospital system nationwide.

An information system should be constructed such that it supports the provision of a nationwide perspective on all ICU beds in Israel so as to divert patients in emergencies where there is an unusual load while simultaneously taking advantage of the relatively short geographical distances in the country.

A multiyear program for intensive care training for medical and nursing teams should be developed, while finding solutions that enable maintaining an adequate professional level even for teams that have undergone intensive care training but do not regularly work in ICUs.

**Conclusions**

Since this study is a noninterventional observational study, the correlations we found (ie, an increase in the percent mortality of ventilated patients with an increase in the census of ventilated patients) are concerning, and suggest, but do not prove, causality. We speculate that the results of this study might help health policy makers to address medical capacities on a nationwide scale, determine how much their system has been affected by pandemics, and to what extent it should be strengthened by the addition or expansion of intensive care facilities. This should facilitate better preparation for future pandemics, the appearance of the next mutant of SARS-CoV-2 that may be resistant to currently available vaccines, or future events potentially resulting in a burden on the health care system. Obviously, there should be a balance between the needs of a country under extraordinary circumstances and needs during “routine” circumstances. An increase in the number of beds might not be the only way out, especially when there are no trained personnel to take care of patients. Increasing ICU training and ICU rotations of health care workers might better prepare a country to handle a large-scale catastrophe.

**Data Availability**

Data are available on the Ministry of Health’s World of Data’s website [32].

**Authors’ Contributions**

JM designed the study and drafted the manuscript. FM conducted the statistical analysis and drafted the manuscript.

**Conflicts of Interest**

None declared.

**References**

https://www.i-jmr.org/2024/1/e41749

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020. URL: [https://tinyurl.com/y82h7bza](https://tinyurl.com/y82h7bza) [accessed 2021-02-25]


Abbreviations

ICU: intensive care unit
LOS: length of stay
MOH: Ministry of Health
OECD: Organisation for Economic Co-operation and Development

©Joseph Mendlovic, Francis B Mimouni. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 09.07.2024. 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.
Chinese Health Insurance in the Digital Era: Bibliometric Study

Zhiyuan Hu¹, BSc; Xiaoping Qin¹, MPH; Kaiyan Chen², BMedSci; Yu-Ni Huang³, PhD; Richard Szewei Wang⁴, MSc; Tao-Hsin Tung⁵, PhD; Yen-Ching Chuang⁶, PhD; Bing-Long Wang¹, PhD, MBA

¹School of Health Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China
²Department of Education, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China
³College of Medical and Health Science, Asia University, Taichung, Taiwan
⁴Affiliation Program of Data Analytics and Business Computing, Stern School of Business, New York University, New York, NY, United States
⁵Evidence-based Medicine Center, Taizhou Hospital of Zhejiang Province affiliated to Wenzhou Medical University, Taizhou, China
⁶Business College, Taizhou University, Taizhou, China

Corresponding Author:
Bing-Long Wang, PhD, MBA
School of Health Policy and Management
Chinese Academy of Medical Sciences & Peking Union Medical College
11th Floor, Block C, Mingyang International Center
No. 46 Xizongbu Hutong, Dongcheng District
Beijing, 100730
China
Phone: 86 15021345074
Email: wangbinglong@sph.pumc.edu.cn

Abstract

Background: China has entered the era of digital health care after years of reforms in the health care system. The use of digital technologies in healthcare services is rapidly increasing, indicating the onset of a new period. The reform of health insurance has also entered a new phase.

Objective: This study aims to investigate the evolution of health care insurance within the context of telemedicine and Internet Plus Healthcare (IPHC) during the digital health care era by using scientometric methods to analyze publication patterns, influential keywords, and research hot spots. It seeks to understand how health care insurance has adapted to the growing integration of IPHC and telemedicine in health care services and the implications for policy and practice.

Methods: A total of 411 high-quality studies were curated from the China National Knowledge Infrastructure (CNKI) database in the Chinese language, scientometric analysis was conducted, and VOSviewer software was used to conduct a visualized analysis of keywords and hot spots in the literature.

Results: The number of articles in this field has increased notably from 2000 to 2022 and has increased annually based on a curve of y=0.332exp(0.4002x) with R²=0.6788. In total, 62 institutions and 811 authors have published research articles in the Chinese language. This study included 290 keywords and formulated a total of 5 hot-topic clusters of “telemedicine,” “IPHC,” “internet hospital,” “health insurance payments,” and “health insurance system.”

Conclusions: Studies on the application of digital technologies in health care insurance has evolved from foundational studies to a broader scope. The emergence of internet hospitals has showcased the potential for integrating IPHC services into insurance payment systems. However, this development also highlights the necessity for enhanced interregional coordination mechanisms. The reform of health insurance payment is contingent upon ongoing advancements in digital technology and increased investment in electronic medical records and primary health care services. Future efforts should focus on integrating technology with administrative systems, advancing mobile health care solutions, and ensuring interoperability among various payment systems to improve efficiency and standardize health care services.


KEYWORDS

telemedicine; health insurance; internet plus healthcare; bibliometric; VOSviewer
Introduction

The Chinese government promulgated a policy document regarding the health insurance system in the last few days of the year in 1998 [1,2] and is currently working on establishing the urban employee’s basic health insurance system, which was initiated in 1999 and completed by the end of 1999, and amendment work has continued for several years thereafter [3]. Since then, China’s current concept of a health insurance system has been established.

During the COVID-19 pandemic, the Chinese government has published numerous policies regarding Internet Plus Healthcare (IPHC)–related health insurance services. The policies are mostly guidance documents related to health insurance payment for IPHC services, including web-based health services and telemedicine health services.

A document aiming at promoting the payment of IPHC in health insurance [4] released by China National Healthcare Security Administration in October 2020 indicated that local governments should design and manage the signing of health insurance agreements for IPHC services. Other official arrangements include improving health insurance payment policies, expanding pilot projects, handling health insurance management, and strengthening supervision measures for newly included health insurance care services among other measures.

As mentioned in a long-term planning official document published by the government of China in December 2022, the importance of IPHC pricing and implementing appropriate services into the health insurance payment list were addressed [5].

IPHC is a novel application of the internet in the health care industry, which includes health education, medical information queries, web-based disease consultations, electronic prescriptions, remote consultations, and various remote forms of health care services such as treatment and rehabilitation [6]. In China, IPHC is an emerging health service model with a cross-industry integration and application of ITs, such as mobile internet, cloud computing, big data, and artificial intelligence [7].

Hence, this field of study is advanced and has real-world implications. The number of studies related to IPHC and telemedicine in 2020, especially in China, has considerably increased [8]. For example, according to reports in early 2020, the number of registrations and IPHC and telemedicine users exponentially increased early during the COVID-19 pandemic; in particular, an internet-based health care platform named “WeDoctor” recorded nearly 80 million visits in early February 2020 and offered services nearly 1 million times. Furthermore, “Ping An Good Doctor,” another major internet-based health care platform claimed to have received over 1 billion visits, and the number of new users has been increasing by several folds [9].

From a historical perspective, telemedicine is an early prototype of China’s IPHC services, which was initiated in the 1990s when doctors in China began communicating with medical experts in other countries through emails about complex and difficult clinical cases. After that, with the increasing use of computers and telecommunications for remote medical consultations in various places, China’s National Healthcare Commission issued regulations to specify the order and behavior of medical care in 1999 to regulate medical order and behavior and enable the development of health care and orderly telemedicine consultation work [6].

In addition, the Chinese government has issued the “Healthy China 2030” project in 2016 [10], which first clearly stated its attitude regarding IPHC, proposing to standardize and promote telemedicine networks and IPHC services and to innovate the IPHC services model.

However, there is no literature using bibliometrics methods, which encompasses this field of the use of health care insurance in IPHC services and telemedicine, and the descriptive study and analysis described here would potentially provide an overview of this area.

This study has 3 objectives: to observe the development of health insurance in telemedicine and IPHC and its related fields in the digital era by examining the publication patterns and key clusters of influential keywords in Chinese. We analyzed the hot spots extracted from high-quality publications and articles based on a bibliometric methodology. We also linked them with future comprehensive studies to illustrate the research frontiers and future roadmap of Chinese health insurance in telemedicine and IPHC service enhancement in the digital era.

Methods

Overview

The bibliometric methodology used in this study describes the landscape and core topics of research in the field from a perspective of health insurance in the advancement in digital health care in China from 2000 to 2022.

Bibliometrics is a method of information analysis, which measures research trends and knowledge structures in a field of research to obtain quantifiable, objective data [11]. The method has been extensively used to quantitatively analyze academic literature to describe trending topics and contributions of scholars, journals, and countries and help researchers understand the current research trends, distribution, and core topics in a given field [12,13]. VOSviewer has better visualization in network and cluster analysis than other software, and the scientometric graphs conform better to current academic research styles. VOSviewer was developed by Nees Jan van Eck and Ludo Waltman and features a powerful bibliometric maps function that can clearly visualize the network of literature, keywords, authors, etc [14]. Using VOSviewer, we generated diagrams for institutional cooperation, keyword co-occurrence, author cooperation, and author cocitation, and the Chinese-language data are all retrieved from the China National Knowledge Infrastructure (CNKI) database [15].

Sampling

All articles related to the fields of health insurance in telemedicine and IPHC published from 2000 to 2022 and written in Chinese are included (Multimedia Appendix 1). The reason
why the beginning year was set as 2000 is that China did not formally establish its current concept of a health insurance system until 1999 [1,2].

We set the CNKI as the target database and retrieved data from “Chinese Journal Full-text Database” and “Academic Journals” (excluding dissertations, conference proceedings, and newspapers). On April 5, 2023, we selected the “Advanced Search” feature and set the search strategy as follows: “(Topic: telemedicine (exact) OR Topic: internet plus healthcare (exact)) AND (Topic: health insurance (exact) OR Topic: medical security (exact)) AND (year range 2000-2022)” to select studies from 2000 to 2022; this yielded 659 articles.

A total of 659 articles retrieved from the search were imported into Excel (Microsoft Corp) for manual checking, and we excluded 6 duplicates papers and then 238 publications belonging to other categories (such as press releases, editorial comments and short reports, nonacademic articles) and 4 publications outside of the time range of 2000-2022 to obtain a final selection of 411 articles; the refinement process is shown in Figure 1. These 411 entries were manually checked to ensure correspondence between the authors and their affiliations, especially when multiple authors are affiliated with the same institution—this step is crucial as it helps avoid potential errors.

We then exported the data and imported them into VOSviewer software (version 1.6.19) for cluster analysis. Based on cluster results, we analyzed and summarized the articles.

Figure 1. The flowchart for data collection. CNKI: China National Knowledge Infrastructure.

Results

Publication Trends

Based on the publication year in the research literature, it can be observed that between 2000 and 2015, the number of articles published in the field of IPHC, remote medical services, and health insurance services was <10 each year, and the accelerated growth started in 2015. Since the number of publications approached 14 in 2015, from 2017 to 2019, the number of articles published in this field plateaued around 30 to 40. In 2020, the number of publications in this research field increased rapidly to nearly 100 papers (around 90 a year) and remain high at about 80-90 papers in 2020 and 2021. Index regression predicts that there will still be a number of Chinese studies published in this field in the near future, while the $R^2$ value of the regression model is 0.6788, indicating that the curve explains the variables relatively well.

Analysis of the Journals

Table 1 shows that journals whose scope includes health insurance focus more on this research field; China Health Insurance Journal accounts for 6.8% of the proportion of studies in this field. Second, journals that explore digital medicine take the lead. China Digital Medicine accounts for 4.8% of studies in this regard, followed by other general medical and health policy research journals. Regarding the distribution, core journals included in Chinese core journals indexed by Peking University or the Chinese Social Sciences Citation Index, both of which are top core collections of Chinese-language journals, enjoy widespread prevalence. The top 5 journals are core Chinese-language journals.
Table 1. The number of articles in the top 10 journals and their proportions in a total of 411 publications.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Journal</th>
<th>Articles, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>China Health Insurance</td>
<td>28 (6.8)</td>
</tr>
<tr>
<td>2</td>
<td>China Digital Medicine</td>
<td>20 (4.8)</td>
</tr>
<tr>
<td>3</td>
<td>Health Economics Research</td>
<td>17 (4.1)</td>
</tr>
<tr>
<td>4</td>
<td>Chinese Hospitals</td>
<td>15 (3.6)</td>
</tr>
<tr>
<td>5</td>
<td>China Social Security</td>
<td>14 (3.4)</td>
</tr>
<tr>
<td>6</td>
<td>Journal of Medical Informatics</td>
<td>11 (2.6)</td>
</tr>
<tr>
<td>7</td>
<td>Chinese Journal of Health Informatics and Management</td>
<td>11 (2.6)</td>
</tr>
<tr>
<td>8</td>
<td>China Health</td>
<td>9 (2.1)</td>
</tr>
<tr>
<td>9</td>
<td>Modern Hospital</td>
<td>8 (1.9)</td>
</tr>
<tr>
<td>10</td>
<td>Chinese Health Service Management</td>
<td>8 (1.9)</td>
</tr>
</tbody>
</table>

Analysis of the Number of Citations of Articles

This field is in its infancy; hence, the top cited article in this field is a study analyzing the basis of this field. The article containing definitions came first, occupying the forefront, received 201 citations (Table 2). Then, latter research focused on currently established modules, problems faced, and future development trends. They all attempted to establish theories and models needed in this field systematically.

Table 2. The top 10 cited Chinese-language articles according to the China National Knowledge Infrastructure.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Title</th>
<th>Journal</th>
<th>Authors</th>
<th>Citations, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Internet + Medical Mode: Contents and System Architecture</td>
<td>Chinese Hospital Management</td>
<td>Zhu Jinsong</td>
<td>201</td>
</tr>
<tr>
<td>2</td>
<td>Research on Development Policy of Integration of Medical Care and Pension and Institutional Pension Service for the Elderly</td>
<td>Medicine and Society</td>
<td>Ma Lili, Chen Na, and Tang Shaoliang</td>
<td>194</td>
</tr>
<tr>
<td>3</td>
<td>The Status Quo of Internet Medical-Based On The Analysis And Investigation on Three Hospitals</td>
<td>Chinese Journal of Health Policy</td>
<td>Wang Anqi and Zheng Xueqian</td>
<td>138</td>
</tr>
<tr>
<td>5</td>
<td>Practice and Exploration of Medical Association in Remote Areas</td>
<td>Modern Hospital Management</td>
<td>Sun Xizhuo, Gong Fangfang, Gu Xiaodong, Su Qian, and Cai Yutong</td>
<td>55</td>
</tr>
<tr>
<td>6</td>
<td>Problems and Countermeasures for “Internet + Healthcare” in China</td>
<td>Administration Reform</td>
<td>Luan Yunbo and Tian Zhendu</td>
<td>52</td>
</tr>
<tr>
<td>7</td>
<td>Problems and Countermeasures for Medical Service Supply in Elderly Care Institutions from the Perspective of Medical-Old-Age Combination</td>
<td>Chinese Journal of Gerontology</td>
<td>Fan Qingmei, Chen Le, Wu Meng, Wu Jiankang, and Li Jiamin</td>
<td>45</td>
</tr>
<tr>
<td>8</td>
<td>Analysis on Problems and Countermeasures of Mobile Health Service in China</td>
<td>Medicine and Philosophy</td>
<td>Yang Xiaoli and Feng Xinwei</td>
<td>45</td>
</tr>
<tr>
<td>9</td>
<td>Study on Regulation System and Related Mechanism of Internet Based Medicine</td>
<td>Chinese Journal of Health Informatics and Management</td>
<td>Meng Qun, Yin Xin, and Dong Kenan</td>
<td>44</td>
</tr>
<tr>
<td>10</td>
<td>Analysis on Service Mode and Application Status for Network Medical Treatment in China</td>
<td>Chinese Journal of Health Informatics and Management</td>
<td>Liu Ning and Chen Min</td>
<td>42</td>
</tr>
</tbody>
</table>

Analysis of Authors

Analysis of the author cooperation network revealed that 811 authors had explored health insurance in IPHC and telemedicine, of whom 9 have acquired more than 12 total link strengths (Figure 2), namely total link strength (TLS), reflecting the strength of cooperation in bibliographical analysis (Table 3). Xu Hong and Lyu Dawei, who discussed the prospect of IPHC in cooperative development of the Changjiang river delta, are the most active authors in this field. The latter active authors also gained over 10 TLSs.
Figure 2. Coauthorship analysis of authors.

Table 3. Top Chinese authors ranked according to total link strength.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Author</th>
<th>Total link strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Xu Hong</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>Lyu Dawei</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>Zheng Xueqian</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>Luo Li</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>Wang Weijun</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>Liu Qian</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>Gong Fangfang</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>Sun Xizhuo</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>Liang Zhigang</td>
<td>12</td>
</tr>
</tbody>
</table>

Analysis of Institutions

A total of 62 institutions were finally included, with a minimum limitation of more than 3 publications, whose publications were analyzed using VOSviewer (Figure 3). Moreover, the School of Health Policy & Management, Nanjing Medical University (TLS=6 times) and other 10 institutions were the top institutions with highest TLS in VOSviewer counting (Table 4). Total link strength and institution co-occurrence of publications.
Figure 3. Institution co-occurrence of publications.

Table 4. Total link strength and institution co-occurrence of publications.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Institution</th>
<th>Total link strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>School of Health Policy &amp; Management, Nanjing Medical University</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Chinese Hospital Association</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Peking University Health Science Center</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Department of Urology, Cancer Hospital Chinese Academy of Medical Sciences</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Beijing Municipal Health Commission</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Nantong University Medical School</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>The Second Hospital of Dalian Medical University, Department of Neurosurgery</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>Institute of Healthy Jiangsu Development, Nanjing Medical University</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>Shanghai Municipal Healthcare Security Bureau</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>Shanghai Institute of Infectious Disease and Biosecurity, School of Public Health of Fudan University</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>Chinese Hospital Association Medical Legality Specialized Committee</td>
<td>6</td>
</tr>
</tbody>
</table>

Analysis of Co-Occurrence of Keywords in Chinese

Identifying trending research fields and directions through keyword co-occurrence analysis is an important indicator for monitoring the development of a discipline. Mapping of keywords is shown in Figure 4, where the size of the node represents the frequency of keyword occurrence, and the lines between the nodes reflect the co-occurrence relationships among multiple keywords. According to the mapping image generated by VOSviewer, current hot topics in Chinese literature in this field can be visually described.

This study analyzed 290 Chinese-language keywords that appeared at least 10 times across included publications using VOSviewer. The results were grouped into 5 clusters: “Telemedicine,” “Internet hospital,” “Internet Plus Healthcare (IPHC),” “Health Insurance Payment,” and “Health Insurance system.” These clusters provide insight into the most prominent
topics related to the use of health insurance in IPHC and telemedicine (Table 5).

**Table 5.** Main keywords translated to English and their total link strength (TLS).

<table>
<thead>
<tr>
<th>Keywords</th>
<th>TLS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster 1: telemedicine</strong></td>
<td></td>
</tr>
<tr>
<td>hierarchical diagnosis and treatment</td>
<td>109</td>
</tr>
<tr>
<td>telemedicine</td>
<td>102</td>
</tr>
<tr>
<td>healthcare alliance</td>
<td>81</td>
</tr>
<tr>
<td><strong>Cluster 2: IPHC</strong></td>
<td></td>
</tr>
<tr>
<td>internet healthcare</td>
<td>129</td>
</tr>
<tr>
<td>healthcare service</td>
<td>68</td>
</tr>
<tr>
<td>healthcare insurance fund</td>
<td>79</td>
</tr>
<tr>
<td><strong>Cluster 3: internet hospital</strong></td>
<td></td>
</tr>
<tr>
<td>internet hospital</td>
<td>179</td>
</tr>
<tr>
<td>smart healthcare insurance</td>
<td>29</td>
</tr>
<tr>
<td>big data</td>
<td>56</td>
</tr>
<tr>
<td><strong>Cluster 4: insurance payment</strong></td>
<td></td>
</tr>
<tr>
<td>online payment</td>
<td>72</td>
</tr>
<tr>
<td>healthcare</td>
<td>31</td>
</tr>
<tr>
<td>smart healthcare</td>
<td>29</td>
</tr>
<tr>
<td><strong>Cluster 5: health insurance</strong></td>
<td></td>
</tr>
<tr>
<td>national health security administration</td>
<td>41</td>
</tr>
<tr>
<td>insured individual</td>
<td>53</td>
</tr>
<tr>
<td>social security card</td>
<td>41</td>
</tr>
</tbody>
</table>

*IPHC: Internet Plus Healthcare.*
Primary Findings From the Co-Occurrence of Keywords
Based on the 5 keyword clusters, the hot spots in the field are described below.

Cluster 1
The primary keyword is “telemedicine” and also includes “remote consultation,” “hierarchical diagnosis and treatment,” “healthcare alliance,” etc, focusing on doctor-to-doctor telemedicine services. Also, it includes topics such as the interactions among different health care service providers under remote conditions, remote pathological analysis, remote consultation, and other telemedicine services based on new-style intelligent communication technology [16]. There are more stakeholders in this type of health care service, demanding that health insurance policies be more detailed to take into account real-world situations.

Cluster 2
The primary keyword is “Internet plus healthcare” (IPHC). Also, it includes keywords such as “Internet healthcare,” “Healthcare service,” “health management,” “health insurance reimbursement,” and “health insurance fund.” In 2019, the National Healthcare Security Administration officially launched the construction of a national unified health care security digitalization platform [17]. In 2021, the system was gradually implemented, and the number of designated medical institutions covered by the basic medical cross-provincial and interregional settlement insurance platform is increasing [17]. The construction of digitalized hospitals has better promoted the development of internet health insurance operations.

Cluster 3
The primary keyword is “internet hospital.” Also, it includes keywords such as “healthcare service,” “online healthcare,” “smart healthcare insurance,” “electronic prescriptions,” and “smart healthcare.” It mainly focuses on providing health management services for patients with chronic disease on the web, especially under lockdown policies for epidemic prevention and control [18]. Furthermore, as an impact of the COVID-19 epidemic, health care service units in various regions have embraced the internet and provided web-based consultation and diagnostic services. Some researchers also involved health insurance payment [17]. In the future, similar methods can be used to provide older adult-focused health care services for the aging population [19].

Cluster 4
The primary keyword is “health insurance payments,” and it also includes “online payments,” “health insurance payment reform,” “insured population,” and “health insurance files.” The research and analyses are mainly policy-oriented, focusing on reducing the burden of health insurance funds and developing reasonable health care service prices and the comparison of different actual implementations of web-based insurance payments in different provinces and municipalities, as well as the future development of smart insurance [20].

Cluster 5
The primary keyword is “health insurance system” and involves keywords such as “medical service prices,” “health insurance funds,” and keywords such as “National Healthcare Security Administration,” and “designated hospitals for health insurance.” The studies and analyses mainly focus on the systematic constructions for health insurance services, the modernization and innovation of the health insurance system from a macro view in the new digital era [21], and the necessary adjustments and changes required for the system to adapt to the new pattern of health care services landscape through the digital era.

Discussion
Principal Findings
Our study shows that in the digital era, China’s health care service system is facing the need for payment mechanisms and policy adjustments to support and optimize the hierarchical health care system. The practice of internet hospitals, as a part of IPHC, has demonstrated the potential for insurance payment in web-based health care services, but it has also revealed challenges in regional integration and interregional coordination. Insurance payment reform, as a key lever for driving systemic change, relies on the advancement of digitalization and informatization, as well as continuous investment in electronic medical records, IT, and primary health care services. Future research and policy making must focus on addressing the integration of technologies with administrative systems, promoting the development of mobile health care, and exploring the interoperability between health care insurance payment systems to achieve efficiency and standardization in health care services.

According to our results, there has been significant development in areas of health insurance in IPHC and telemedicine research over the past 2 decades. The number of relevant publications has steadily increased year by year, and more than 60 researching institutions and over 800 authors in China have published academic research papers in this field. Since 2017, the number of publications in this field has risen sharply from around 30 papers each year to around 90 papers each year, which means that in the near future, an increasing number of studies will focus on the use of health insurance in IPHC and telemedicine, especially in improving health insurance implementational methods and measures, and future patching policies regarding digital health care services.

In the founding stage of the research field, the most highly cited paper was an analysis of its theoretical basis, with 201 citations, occupying the top spot, having focused on established modules, current challenges, and future development trends, striving to systematically develop theories and models required for the field.

Among 62 institutions, over 10 of them had a TLS of 6 and rank at the top.

There were 62 institutions with a minimum of >3 publications; among them, 11 institutions with 6 TLSs were the top academic institutions. These research institutions include universities, hospitals, and associations in the health care industry, implying
that this research field has received widespread attention in Chinese academic fields and reflects the importance of research in this field in China. Different types of institutions engage in academic discussions from their specific perspectives.

Through bibliometric and visualization analysis, we gained a deeper understanding of the overall landscape of this research field, including prominent Chinese authors and publishing institutions, as well as their collaborative relationships and academic influence. This information provides researchers with transparent channels for selectively obtaining advanced and valuable research results. Co-occurrence analysis can also depict research trends and hot spots [22], providing researchers with assistance in proposing research topics to convince funding agencies to develop more effective funding plans.

Analysis of Research Focus

Overview

Keywords are essential in a research article and contain the most important information [23].

Based on the analysis of keyword clusters in the literature [24], this study focuses on different aspects and levels of research directions that health insurance needs to promote in implementing IPHC, as well as the problems and possible solutions reflected in the actual health care service practice [25]. From the integration of health care services and medical treatment to the connection between health insurance systems, based on the interconnectedness of information platforms [26], the key conflicts encountered in health care service practice are gradually being resolved and improved [27]. There are a total of 5 clusters, and based on the research results, the main research frontiers involved in these 5 clusters will be discussed below.

Cluster 1

The primary keyword is “telemedicine.” By constructing a telemedicine service system between institutions in a medical treatment combination, which involves multiparty participation such as bidirectional referral and web-based consultation [28], the mechanism of health insurance payment needs to be strengthened and improved. Furthermore, to improve the hierarchical health care system, a mixed health insurance payment model should be explored [29]. In the context of combining remote health care services with health insurance, the current emerging problems such as the barriers to the generation of cross-provincial systems need to be clearly defined, and policies need to be adjusted and improved accordingly [30].

Cluster 2

The primary keyword is “Internet plus healthcare.” The development of IPHC requires continuous advancement in medical digitalization [31], strong modern digital technologies such as artificial intelligence, digital twins, big data management, and remote services to realize many health care services that are currently only in their infancy, such as smart health care, which are the frontiers of current research [32]. The platform construction of electronic medical records and electronic prescriptions has been integrated with intelligent health care security platforms. However, in certain specific implementation, there are still many practical issues such as technology advancing ahead of management models [33]. Advancing creatively through a combination of technology and policies is necessary to solve many potential new problems and conflicts. In the future, IPHC will remain an important theme for China’s medical and health reforms and development [34], and there will inevitably be more academic studies on policies and management for this topic.

Cluster 3

The primary keyword is “internet hospital.” Some internet hospitals have connected with local health insurance individual account payment channels to achieve health insurance reimbursement [35]. The policy enables follow-up services for chronic diseases to be included in the scope of health insurance payment, accomplishing a series of closely matched services such as web-based remote consultation [19], web-based prescription, and health insurance payment for purchasing medication [18]. Health management can also be beneficial, such as improving medication compliance and strengthening awareness of chronic disease treatment and community health management [36]. The inclusion of internet follow-up services in health insurance payment emphasizes the homogenization of offline and web-based diagnosis and treatment behaviors [37], providing a basis for health insurance pricing through the use of advanced technological assessments and other means [20], and requires further research, which is a topic of high interest.

Cluster 4

The primary keyword is “health insurance payments.” The use of health insurance in internet-based medical payment is increasing, but problems and difficulties have emerged. For example, there are significant problems in the integration of regional health insurance, internet-based medical platforms, and local health insurance systems, as well as communication and coordination issues such as cross-regional medical treatment and settlement [38]. As one of the main directions for future development of internet-based health care services, the use of mobile health care through smartphones is also a current research hot spot [39].

Cluster 5

As the most powerful lever to drive reforms in the entire medical system, health care payment reforms are crucially supported by digital and information-based payment systems, so the IPHC with health insurance occupies an important position [40]. “Electronic medical records” for referral purposes and more sophisticated IT can accelerate the implementation of payment reforms [41]. Vigorous development of primary care–based internet health care [42], increasing the proportion of health insurance expenditures in this area, among other measures. Research frontiers include problems with interoperability among internet health care systems, hospital internal management platforms [43], electronic medical record systems, imaging and inspection platforms [44], and other related issues.

Strength and Limitations

To the best of our knowledge, no study has carried out bibliometric analysis in this field of research on the use of health
insurance in IPHC services, along with telemedicine. It is of noticeable significance to discuss these issues and it is now necessary for us to study the research topic and identify hot spots. The method of bibliometrics and visual analyses enable us to sort out research focuses in recent publications as well as their correlation and differentiation.

However, this study inevitably has some limitations. We only retrieved studies and reviews on research topics related to this field from the CNKI. Although the CNKI plays a significant role in academic research and literature analysis in China, there are some obvious limitations to its application in international research fields, and studies need to consider these limitations and comprehensively integrate database resources in accordance with the specific needs of the research, to ensure comprehensive evaluation and analysis. Therefore, we may have missed some publications due to database limitation, and articles related to other languages may not have been included.

Bibliometric methods provide an overall insight into the landscape of a specific research field, but researchers and policy makers should be aware that the feature is not detailed enough for the evaluation or decision-making, and this study provides no enough in-depth insight into the influential articles and authors in this field. Balanced approaches that integrate bibliometrics research with other assessment methods can provide a thorough understanding of research impact and trends.

Future Directions

Recently, many previously ignored issues have been discussed and are now at the forefront, and numerous real-world problems associated with internet-based health care services, which has been limited by the scale of services, have been pointed out and discussed. Starting from the postpandemic era, the demand for remote or IPHC services will continue to grow. The deepening support of policies for IPHC services can meet the demand for health insurance management, mobile insurance payment, and cost reduction. Future research hot spots are developing, such as those focusing on the web-based application of live broadcasting, new media, artificial intelligence technology, etc, into IPHC services. Other hot spots include chronic disease management and primary health care, as well as the community older adult care in health insurance reforms in the IPHC era. These hot spots are surely important for suggesting more reasonable policy measures, enhancing the accessibility of health care services, reducing costs, and improving the quality of health care, thus better serving the people requiring them.

In summary, this study on the application of health insurance in internet-based health care services is quite forward-looking, and it is an important frontier for future health care service research.

Significance

This study describes a bibliometric analysis of the current high-quality Chinese literature on the application of telemedicine and IPHC in health insurance in China. This study used the popular software tool VOSviewer (version 1.6.19) to analyze the literature published in the CNKI, which was involved in the development of this field, and provides an overview of all the existing high-quality Chinese research and guides future research developments to improve the application of telemedicine and IPHC in health insurance. In particular, the unique IPHC paradigm from China is of importance to health care professionals worldwide.

Conclusions

This study used bibliometric analysis to describe the current situation and trends of health insurance in IPHC and telemedicine from the literature in China. This study highlights prominent research institutions, hospital researchers, and researchers at research institutes in universities engaged in this field. More articles on health insurance in IPHC and telemedicine are expected to be published in the next few years. The use of internet hospitals has underscored the potential of health care insurance payment in IPHC services, but it also highlights challenges pertaining to regional integration and interregional coordination. As the key lever for instigating systemic change, health care insurance payment reforms hinge on the progression of digitalization and informatization, along with ongoing investment in electronic medical records, IT, and primary health care services. Future research and policy formulation are expected to focus on tackling the integration of technologies with administrative systems, fostering the advancement of mobile health care, and delving into the interoperability among health care payment systems to attain efficiency and standardization in medical health care. This research provides an invaluable reference, enhancing our grasp of the current landscape and prospective progress in the field of health insurance within the domains of IPHC and telemedicine.

Acknowledgments

The authors thank the editor-in-chief and the journal reviewers for their helpful comments, which have helped us improve our manuscript significantly. No human or animal studies are reported in this paper, and no potentially identifiable human images or data are presented in this study. This research was supported by the Chinese Academy of Medical Sciences and Peking Union Medical College, China (grant 2021-RC630-001).

Data Availability

All data and material generated or analyzed in this study are included in this published paper.
Authors’ Contributions

ZH, KC, and B-LW designed the study and wrote the manuscript. XQ, RSW, Y-CC, T-HT, and Y-NH critically reviewed the manuscript. ZH, RSW, YQ, and Y-NH directed statistical analysis and helped interpret the results. Y-NH, RSW, Y-CC, T-HT, and B-LW edited the manuscript. All authors reviewed and approved the manuscript. All authors have read and agreed with the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The number of cumulative publications (2000-2022) and model fitting curve.

References


42. Jiang H, Xu A, Dai Z. Study on internet medical information spillovers and health care choice of urban and rural residents—based on the Sixth Health Service Survey data of Jiangsu province. Chinese Primary Health Care 2022;36(03):17-20 [FREE Full text]


**Abbreviations**

CNKI: China National Knowledge Infrastructure

IPHC: Internet Plus Healthcare
TLS: total link strength
Biochemical Changes in Adult Male Gamers During Prolonged Gaming: Pilot Study

Kasper Bygum Krarup¹, BSc, MD; Johannes Riis¹, MD; Morten Mørk², MD, PhD; Hien Thi Thu Nguyen³, MSc, PhD; Inge Søkilde Pedersen³, MSc, PhD; Søren Risom Kristensen², MSc, MD, PhD; Aase Handberg², MSc, PhD; Henrik Bygum Krarup², MSc, PhD

¹Department of Geriatrics, Aalborg University Hospital, Aalborg, Denmark
²Department of Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark
³Department of Molecular Diagnostics, Aalborg University Hospital, Aalborg, Denmark
⁴Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

Corresponding Author:
Kasper Bygum Krarup, BSc, MD
Department of Geriatrics
Aalborg University Hospital
Reberbensgade 15
Aalborg, 9000
Denmark
Phone: 45 61781289
Email: kasper.krarup@rn.dk

Abstract

Background: Gaming has become an integrated part of life for children and adults worldwide. Previous studies on the impact of gaming on biochemical parameters have primarily addressed the acute effects of gaming. The literature is limited, and the study designs are very diverse. The parameters that have been investigated most thoroughly are blood glucose and cortisol.

Objective: This exploratory study is the first to investigate the effects of long gaming sessions on the biochemical parameters of healthy male adults. The extensive testing allowed us to observe short-term changes (within 6 hours), long-term changes during the duration of the gaming sessions, and follow-up after 1 week to determine whether any changes were longer lasting.

Methods: In total, 9 experienced gamers completed 2 back-to-back 18-hour gaming sessions interspersed with a 6-hour rest period. All participants adhered to a structured sleep pattern due to daytime employment or attending university. Blood, saliva, and urine samples were collected from the participants every 6 hours. Linear mixed-effect models were used to analyze the repeated-measures data accumulated during the study. A total of 51 biochemical parameters were investigated.

Results: In total, 12 of the 51 biochemical parameters significantly changed during the study: alkaline phosphatase, aspartate aminotransferase, bilirubin, chloride, creatinine, glucose, hemoglobin, immature reticulocyte fraction, lactate, methemoglobin, sodium, and thrombocytes. All changes were within the normal range. The mean glucose level of the participants was 4.39 (SD 0.07) mmol/L at baseline, which increased significantly by 0.24 (SD 0.07) mmol/L per 6 hours during the first period and by 0.38 (SD 0.07) mmol/L per 6 hours in the second period (p < .001). The glucose levels during the second session increased even though the participants had little energy intake. Cortisol levels did not change significantly, although the cortisol pattern deviated from the typical circadian rhythm. During both gaming sessions, we observed increasing cortisol levels from 6 AM until noon. The participants were relatively dehydrated at the start of the study. The patients were asked to fast before the first blood sampling. Within the first 6 hours of the study, the participants rehydrated, followed by relative dehydration during the remainder of the study. This pattern was identified using the following parameters: albumin, creatinine, hemoglobin, erythrocytes, potassium, and platelets.

Conclusions: This study is the first of its kind, and many of the analyses in the study yielded novel results. The study was designed to emulate the behavior of gamers during the weekend and other long gaming sessions. At this point, we are not able to determine the difference between the effects of gaming and behavior during gaming. Regardless, the results of this study suggest that healthy gamers can partake in long gaming sessions, with ample amounts of unhealthy foods and little rest, without acute impacts on health.
Introduction

Video games have become a favorite pastime among children and adolescents. In the United States, 99% and 94% of boys and girls, respectively, play video games [1]. Video games are part of a larger category of sedentary activities linked to health issues such as physical inactivity, overeating, obesity, and diabetes [2-6]. Collegiate-level gamers have a comparable BMI to their nongaming peers but are less active with a higher body fat percentage, lower lean body mass, and lower bone mineral content [7]. This is particularly troublesome because studies report that children spend as much as 7-11 hours daily engaged in screen-based activities after school [8]. In addition, gaming has been linked to overeating and overconsumption of soft drinks as well as adverse health behaviors [9-12]. In a recent review, the authors found that while gaming may increase energy expenditure above baseline levels, gaming does not constitute physical activity. Energy expenditure may increase, but the activity level is comparable to that of standing or walking [13].

Previously published data from this study showed that the participants ingested an excessive number of calories from both food and drink [14]. During the study, the participants ingested an average of 6160 kcal from food and 1844 kcal from liquid sources. Additionally, the participants consumed 1354 mg of caffeine on average during the same period [14].

Previous studies of the impact of gaming on biochemical parameters have primarily addressed the acute effects of gaming. The literature is limited, and the study designs are very diverse. The parameters that have been investigated most thoroughly are blood glucose and cortisol. For blood glucose levels, gaming does not appear to increase glucose levels within the first 20 minutes of a gaming session [15,16]. Chaput et al [15] reported that gaming increased blood glucose levels after 40 minutes of a 60-minute gaming session.

Cortisol has been used as a marker of both physiological and mental stress during gaming [17-19]. Gaming may affect cortisol levels, but the nature of this relationship has yet to be elucidated. Oxford et al [17] reported that cortisol levels increase acutely when gamers compete against friends. The authors suggested that this was typical of male-male competition behavior [17].

While biochemical markers such as blood glucose and cortisol levels have been investigated during the last 2 decades, many biochemical markers pertaining to acute changes in the health of the human body have yet to be investigated. In this study, we aimed to investigate a broad array of biochemical markers to assess homeostasis, lipid metabolism, internal organ function, hematological balance, acid-base balance, and blood gases during long gaming sessions. This is the first time that most of the included parameters (except for glucose and cortisol) were investigated in gamers and during long gaming sessions. As such, we did not know what to expect over the course of the study.

According to a recent literature review, knowledge of the effects of long video gaming sessions is minimal [13]. This exploratory study is the first of its kind in gaming. This is true for both the length of the gaming sessions, the physical setup, and the extensive testing.

A limitation of this study was the relatively small number of participants. This study was designed to realistically emulate the gaming behavior of young adults in a controlled setting in a hospital dining hall and an adjoining meeting room. The study was conducted as a local area network party for practical reasons. Four laboratory technicians worked at all times to sample, prepare, and analyze blood samples. A doctor and 2 investigators were also present throughout the gaming sessions.

We tried to overcome the artificial situation of the event by discussing the setup with the participants before the study to create the most real-life-like experience (video clip [20]). The participants were asked to consume food and drink according to their wishes and what they would habitually consume [14]. The extensive testing in the study allowed us to observe short-term changes (within 6 hours), long-term changes during the duration of the gaming sessions, and follow-up after 1 week to determine whether any changes were longer lasting. The aim of this study was to investigate the effects of long gaming sessions on the biochemical parameters of healthy male adults.

Methods

Participants and Intervention

We have previously presented how the study was conducted and the physiological response to long gaming sessions [14]. In this paper, the results of an extensive collection of biochemical data are presented. We have adhered to the CONSORT (Consolidated Standards of Reporting Trials) statement regarding pilot and feasibility trials [21]. A CONSORT checklist can be found in Multimedia Appendix 1.

In brief, 9 healthy male participants older than 18 years of age with vast gaming experience were enrolled. The mean age of the participants was 25.8 (SD 2.6) years, and the mean BMI was 24.8 (SD 2.9) years. All participants were either full-time students or employees [14]. According to the protocol (EudraCT 2019-004091-19), the plan was to enroll 6-9 gamers, with at least 6 gamers engaging in gaming for 48 hours. Participants were enrolled after they had called for participants from local e-sports clubs through e-sports instructors, online message boards, and word of mouth. The participants were recruited into 2 teams of 4 and 5 members.

The intervention consisted of two 18-hour gaming sessions interspersed with a 6-hour break. During the break, the participants had approximately 4 hours of sleep. After the last...
gaming period, all participants underwent both physiological and biochemical tests.

Throughout the intervention, the participants had ad libitum access to food and drink. Before the study, participants provided lists of their preferred snacks and drinks. An assortment of chips, candies, cookies, buns, cold cuts, cereal, and fruit was available throughout the study. Additionally, an evening meal was provided at 6 PM during both gaming sessions: pizza on the first evening and hamburgers on the second evening. Participants had access to tap water, soda, energy drinks, coffee, milk, and chocolate milk.

The participants were instructed to avoid strenuous physical labor, cardiovascular exercise, alcohol, and junk food for 7 days before the study. Additionally, on the day of the study, the participants were instructed to stop food intake at noon and only drink water in case of thirst.

During the study, the participants were not restricted in any way regarding gaming. Specifically, all types of games across all genres and platforms were allowed. One participant brought his PlayStation, and as a new soccer game had just been released, participants played matches in teams or head-to-head. At other times, the participants played alone in certain games or played on the web with and against players who were not participants in the study. Especially during the last 18 hours of the study, the participants played first-person shooter games against each other on teams.

**Participant Involvement**

After the protocol was approved by the local ethics committee in February 2018, in total, 2 meetings were held with potential participants and local e-sports instructors, discussing the content of the protocol. In particular, food monitoring and blood sampling methods were changed in accordance with the wishes of the participants.

The participants suggested a “food diary” to monitor their caloric intake. This suggestion was incorporated into the protocol. First, it was suggested by the research team that a venous catheter be inserted for blood sampling throughout the study. Instead, the participants opted for multiple venipunctures. The changes suggested at the meetings were approved by the local ethics committee in an amendment before the study started.

**Ethical Considerations**

The North Denmark Region Committee approved the study protocol on Health Research Ethics (N-20180011; EudraCT 2019-004091). Each participant provided informed consent in writing twice, 2 weeks prior to the study and again on the day of the study. The study used a standardized consent form that stated that participation was voluntary and that participants could withdraw from the study at any time without reason or consequence. Primary consent was obtained after approval for the secondary analysis. The data were deidentified. Except for travel expenses, participants were not compensated for their participation.

**Blood Sampling and Processing**

When the participants arrived at the laboratory, baseline samples were collected, including a venous blood sample, a urine sample, and a saliva swab. Blood, urine, and saliva were collected from the participants every 6 hours. After the blood was drawn by venous puncture, it was taken to the laboratory. Biochemical analyses were performed immediately using an ABL800 FLEX Blood Gas Analyzer (Radiometer); other samples were centrifuged (at 3000 rpm/1875xg for 10 minutes), and the plasma was frozen at −80 °C for later analysis. Blood samples were analyzed using a Cobas 8000 Modular Analyzer (Roche Applied Science) and a Sysmex XN-9000 Hematology Analyzer (Sysmex Europe, GmbH). In total, 18 parameters were analyzed on the Cobas 8000 (1 parameter is only presented in Multimedia Appendices 2 and 3), 22 parameters were analyzed on the Sysmex XN-9000 (16 parameters are only presented in Multimedia Appendices 2 and 3), and 18 parameters were analyzed using the ABL800 (11 parameters are only presented in Multimedia Appendices 2 and 3). An overview of the analyses by apparatus can be found in Multimedia Appendix 2. The complete summary of the results is provided in Multimedia Appendix 3. A total of 8 parameters were analyzed twice between the 3 machines (calcium [Ca], glucose, lactate, potassium [K], sodium [Na], bilirubin, creatinine, and hemoglobin [Hb]). All the tests were performed in accordance with the manufacturer’s instructions in a nationally accredited biochemistry laboratory (Department of Clinical Biochemistry, Aalborg University Hospital).

The following 30 parameters were measured and are presented in the Results section: glucose (mmol/L), lactate (mmol/L), cortisol (mmol/L), Ca (mmol/L), albumin-corrected calcium (mmol/L), K (mmol/L), Na (mmol/L), and chloride (Cl, mmol/L) were measured to assess homeostasis. Low-density lipoprotein (LDL) cholesterol (mmol/L), high-density lipoprotein cholesterol (mmol/L), total cholesterol (mmol/L), and triglyceride (TG, mmol/L) levels were measured to assess lipid metabolism. Alanine aminotransferase (ALT, U/L), albumin (g/L), alkaline phosphatase (ALP, U/L), bilirubin (μmol/L), creatinine (μmol/L), C-reactive protein (CRP, mg/L), and ferritin (μg/L) were measured to assess organ-specific markers. S O₂ (%) , pCO₂ (kPa), pO₂ (kPa), pH, and standard bicarbonate concentration (mmol/L) were measured to assess acid-base balance and blood gases. Erythrocytes (10¹²/L), erythrocyte volume fraction (EVF), Hb (mmol/L), mean cell Hb (10⁻¹⁵ , SE 0.0018 mol), mean cell volume (10⁻¹⁵/L, SE 0.088/L), and platelets (10⁹/L) were measured to assess the hematological markers.

**Statistical Analyses**

Linear mixed-effects models were used to analyze participants’ absolute changes in biochemical parameters throughout the study. This modeling method is a standard extension of linear regression models that controls for the random effects introduced by having paired data and is often used when analyzing studies with repeated measurements. We performed separate univariate analyses for each change in each biochemical parameter; time was the fixed effect, and participants were the random effect.
Additionally, we included an interaction term between time and period. This gave separate results for the first and second gaming periods.

We used a likelihood ratio test to test the statistical significance of the linear mixed effects models compared to a basic model, which included only random effects and thereby assumed no development over time. In total, 51 markers were analyzed (58 markers in total and 7 duplicates), and Bonferroni correction was applied by multiplying $\alpha$ (originally .05) by the number of tests. Statistical significance was set at $\alpha=0.05$ before Bonferroni correction [22]. The assumptions of linear mixed-effects models were tested and fulfilled in all the analyses. Linearity and homoscedasticity were assessed by inspecting residual plots. The normality of the residuals was assessed by inspecting histograms and quantile-quantile plots. Given that this was an exploratory study, outliers were not removed. Due to the sample size and number of tests, we did not perform post hoc tests comparing the data at individual time points.

Graphs illustrating changes over time were made for all variables, consisting of average values among participants for each sampling point, together with SEs presented as error bars from the mean. In addition, selected graphs were included to illustrate developments over time for specific parameters (Figures 1-4) that were not necessarily reflected in the linear mixed-effects models.

One participant’s LDL cholesterol level decreased 3-fold below the minimum measuring range. The missing values were substituted by using the previous lowest value of the participant [23]. All the statistical analyses were performed using Microsoft Excel 2013 (Microsoft Corp) and RStudio (version 1.1.383; Posit, PBC). Linear mixed-effects models were generated using the lme4 package [24].

Figure 1. Dehydration. (A) Albumin, (B) hemoglobin, (C) erythrocyte volume fraction (EVF), and (D) platelets.
Figure 2. Glucose development over time.

Figure 3. Cortisol development over time.
**Results**

**Overview**

The linear mixed-effects models produced results with a mean baseline value at the start of the study and coefficients of change for the first and second gaming periods. Several parameters, such as albumin, Hb, EVF, and the platelet count, exhibited a pattern of high initial values, followed by a decrease after 6 hours and a continuous increase during the rest of the gaming period (Figure 1). The complete list of results of all analyses is available in Multimedia Appendix 3.

**Homeostasis**

The mean glucose level of the participants was 4.39 (SD 0.07) mmol/L at baseline, which increased significantly by 0.24 (SD 0.07) mmol/L per 6 hours during the first period and by 0.38 (SD 0.07) mmol/L per 6 hours in the second period (Figure 2). The mean Na level was 141.9 (SE 0.12) mmol/L at baseline and decreased by 0.31 (SD 0.1) mmol/L per 6 hours during the second period. There were no changes during the first period.

The mean Cl level was 105.9 (SE 0.14) mmol/L at baseline and decreased significantly by 0.54 (SD 0.14) mmol/L per 6 hours during the second period, and there were no significant changes during the first period. Lactate, Ca, K, and albumin-corrected calcium levels did not significantly change during the study. Cortisol levels also did not change significantly. However, cortisol levels appeared to be affected by long gaming sessions and a lack of rest during the study. A deviation from this pattern is usually associated with daily fluctuations in cortisol levels (Figure 3). During a normal cortisol rhythm, cortisol levels are at their lowest level at midnight before increasing to their peak value at 6 AM, followed by a steady decline toward midnight.

**Lipid Metabolism**

Total cholesterol, LDL, high-density lipoprotein, and TG did not significantly change during the study. However, the TG levels during the study showed a distinct pattern: TG levels increased after evening meals but returned to baseline after 6 hours (Figure 4).

**Organ-Specific Markers**

The mean ALP level was 77.2 (SE 0.63) U/L at baseline and increased during the second period by 3.8 (SD 0.63) U/L per 6 hours. There was no significant change during the first period. The mean bilirubin level was 9.9 (SE 0.22) μmol/L at baseline, decreased by 1.15 (SD 0.22) μmol/L per 6 hours during the first period and decreased by 1.05 (SD 0.22) μmol/L per 6 hours during the second period. The mean creatinine level was 93.6 (0.56) μmol/L at baseline, decreased by 2.9 (SD 0.56) μmol/L per every 6 hours during the first period and decreased by 1.15 (SD 0.56) μmol/L per every 6 hours during the second period. Albumin, CRP, ferritin, and ALT did not significantly change during the gaming sessions.

**Acid-Base Balance and Blood Gases**

sO₂, pCO₂, pO₂, pH, and standard bicarbonate concentration did not significantly change during the study.

**Hematology**

The mean Hb level was 9.18 (SE 0.025) mmol/L at baseline, and it decreased by 0.055 (SD 0.025) mmol/L every 6 hours during the first period. However, during the second period, Hb
increased by 0.027 (SD 0.025) mmol/L per 6 hours. The mean platelet count was 258×10⁹/L (SE 1.1×10⁹/L) at baseline and increased by 6.9×10⁹/L (SD 1.1×10⁹/L) per 6 hours during the second period. During the first session, there were no significant changes. Erythrocytes, EVF, mean cell Hb, and mean cell volume did not significantly change during the study.

Discussion

Principal Findings
This study is the first of its kind regarding the design and number of biochemical analyses in the gaming population. The fact that the study stretches over 42 hours of continuous measurements makes it unique compared to the current literature on gaming science. The study applies to recreational gamers who play with various levels of seriousness but lack a singular focus on the competition associated with e-sports. Overall, the effect of gaming on standard biochemical parameters in healthy male adults is limited. Significant changes were found in 12 of the 51 parameters. Most of the results of biochemical tests are novel findings in a gamer population [15]. It is not surprising that most parameters were unaffected by the intervention, but the high number of examined parameters added to our understanding of the effects of gaming. The participants had a large intake of calories throughout the gaming sessions (especially during the first gaming sessions) [14]. We found small changes in the levels of several biochemical and hematological biomarkers, but all the levels were within the normal range. Overall, the results of this exploratory study suggest that, from a biochemical and hematological standpoint, the health of male adults is not altered in the short term by long gaming sessions.

Dehydration
The development of several parameters, including albumin, ALP, creatinine, Ca, ferritin, Hb, erythrocytes, EVF, K, and platelets, over time suggested that the participants were relatively dehydrated at the start of the study and rehydrated within the first 6 hours, followed by relative dehydration during the remainder of the study (Figure 1). Dehydration, despite a 3-L fluid intake per 18-hour gaming session, was most likely aided by the participants’ caffeine intake, as a large intake of caffeine during nonstrenuous activities can cause increased diuresis [25].

Homeostasis
Blood glucose levels increased consistently during each of the gaming sessions (Figure 2). This development agrees with the findings of Chaput et al [15], but the changes in this study occurred over a much longer period. The changes in blood glucose levels cannot be attributed to dehydration, as this pathway is under tight hormonal control [26]. The glucose levels during the second session increased even though the participants had little energy intake. This could be due to the unexpectedly high levels of cortisol present at the same time (Figure 3), triggering the release of glucose from body stores. The glucose levels never exceeded the normal range and returned to baseline levels 1 week after the second long gaming session. Short- and long-term changes in blood glucose levels in gamers who regularly participate in long gaming sessions need further investigation.

The development of cortisol during the study differed from what would be expected (Figure 3). Cortisol is typically at its lowest level at midnight and increases sharply at 6 AM to reach a maximum between 6 and 10 AM, after which a sharp decline is expected [27]. The continued increase at noon during both sessions is surprising. Typically, the cortisol concentration decreases throughout the day and evening, reaching its nadir at midnight. The participants slept between noon (after the first session) and 6 PM. The 6 PM cortisol value could have been influenced by the waking cortisol response, which is related to the circadian rhythm [28]. Multiple factors potentially contribute to the disruption of the regular cortisol rhythm. Going into the study, the participants all adhered to a structured sleep pattern due to daytime employment or attending university classes. During both gaming sessions, we observed a rise in cortisol levels from 6 AM until noon. Gaming could be the cause of this change, as the level of alertness (or “stress”) in gaming potentially requires the continuous activity of the hypothalamic-pituitary-adrenal axis [28,29]. An increase in cortisol based on light intensity has been described, and light from monitors could be a factor in the changes observed [30]. Nightshift workers who are habitually awake during the night exhibit markedly lower morning cortisol levels [31]. The sharp decrease in cortisol levels at midnight (both nights) indicated that the regular diurnal slope was not completely changed by the weekend during extreme gaming.

Lipid Metabolism
Gaming and sedentary behavior have been associated with obesity and increased cholesterol and TG levels [4], as they displace other nonsedentary activities [32]. Based on standard biochemistry, cholesterol and TG were unaffected by long gaming sessions and unrestricted food intake. However, TG levels sharply increased after mealtimes, which were normalized 6-12 hours after each meal. This increase was much greater than usual [33,34]. The sharp increase in TG was most likely the result of the high fat content of the ingested foods [14].

Organ-Specific Markers
Overall, this study does not indicate that long gaming sessions impact the kidneys or livers of gamers. Creatinine levels decreased slightly during the study, possibly because of reduced physical activity. A decrease in creatinine at this level is not known to influence health.

Bilirubin and ALP levels decreased significantly during the study. This decrease, while significant, is not associated with any known pathophysiology. Additionally, the large amounts of caffeine ingested may have lowered the bilirubin levels [35]. In experimental models, caffeine has been shown to have antifibrogenic, anti-inflammatory, and antioxidant effects that potentially exert liver protection [36-39].

We expected an increase in ALT based on food consumption during the study [40]. Likewise, we suspected that the inflammatory parameters would change due to the large intake...
of unhealthy food and drinks. As CRP, ferritin, and ALT are sensitive to changes in behavior (strenuous exercise, alcohol consumption, and excessive food intake), their lack of change during the study indicates a minor or no impact of gaming behavior.

**Acid-Base Balance and Blood Gases**

Our results showed that gaming did not affect the pH balance or its regulatory systems. Furthermore, the $P_{O_2}$ in the venous blood did not significantly change throughout the trial. This finding agrees with the literature suggesting that sedentary gaming is not physically demanding [13].

**Hematology**

Hematological parameters have not previously been examined in gamers during long gaming sessions. The Hb concentration decreased significantly during the first 18 hours of gaming and increased during the second 18-hour session. These changes are in accordance with the changes in hydration status outlined earlier. Thrombocyte counts increased throughout the study, especially in the second 18-hour session. Despite the significant changes, the parameters were always within the normal range. Overall, this exploratory study does not suggest a need for further investigation into the association between hematological parameters and gaming.

**Strengths and Limitations**

The participants were all male individuals in their 20s, and there were only 9 participants. Based on these results, it is not possible to determine whether gaming has a different or even harmful effect on children, adolescents, or female individuals. The study was conducted in a hospital and not in the familiar environments of the participants. This may have caused participants to alter their behavior in ways we cannot determine. The easy availability of food and snacks may not relate entirely to how long gaming sessions are conducted at home or at local area network parties. We cannot rule out the possibility that overeating occurred, but the available food and drinks were only present at the behest of the participants. During the first 6-12 hours of the study, overeating was perhaps a factor, but during the second session, the participants had a markedly lower intake of calories from both food and drinks [14]. On the basis of this study, we are not able to determine whether the excessive intake of calories was due to convenience or to sustain a high level of performance during strenuous gaming.

**Conclusions**

This study is the first of its kind, and many of the analyses in the study yielded novel results. The study was designed to emulate the behavior of gamers during the weekend and other long gaming sessions. At this point, we are not able to determine the difference between the effects of gaming and behavior during gaming. Regardless, the results of this study suggest that healthy gamers can partake in long gaming sessions, with ample amounts of unhealthy foods and little rest, without acute impacts on health.
References


Abbreviations

ALP: alkaline phosphatase
ALT: alanine aminotransferase
Ca: calcium
Cl: chloride
CONSORT: Consolidated Standards of Reporting Trials
CRP: C-reactive protein
EVF: erythrocyte volume fraction
Hb: hemoglobin
K: potassium
LDL: low-density lipoprotein
Na: sodium

https://www.i-jmr.org/2024/1/e46570
Interact J Med Res 2024 | vol. 13 | e46570 | p.267
(page number not for citation purposes)
TG: triglyceride
Effectiveness of a Smartphone App to Promote Physical Activity Among Persons With Type 2 Diabetes: Randomized Controlled Trial

Stephanie E Bonn¹, PhD; Madeleine Hummel¹, MD; Giulia Peveri²,³, PhD; Helén Eke¹, MSc; Christina Alexandrou⁴, PhD; Rino Bellocco²,⁵, PhD; Marie Löf⁶, PhD; Ylva Trolle Lagerros¹,⁶, MD, PhD

¹Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
³Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy
⁴Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden
⁵Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
⁶Center for Obesity, Academic Specialist Center, Stockholm Health Services, Stockholm, Sweden

Abstract

Background: Physical activity is well known to have beneficial effects on glycemic control and to reduce risk factors for cardiovascular disease in persons with type 2 diabetes. Yet, successful implementation of lifestyle interventions targeting physical activity in primary care has shown to be difficult. Smartphone apps may provide useful tools to support physical activity. The DiaCert app was specifically designed for integration into primary care and is an automated mobile health (mHealth) solution promoting daily walking.

Objective: This study aimed to investigate the effect of a 3-month-long intervention promoting physical activity through the use of the DiaCert app among persons with type 2 diabetes in Sweden. Our primary objective was to assess the effect on moderate to vigorous physical activity (MVPA) at 3 months of follow-up. Our secondary objective was to assess the effect on MVPA at 6 months of follow-up and on BMI, waist circumference, hemoglobin A₁c, blood lipids, and blood pressure at 3 and 6 months of follow-up.

Methods: We recruited men and women with type 2 diabetes from 5 primary health care centers and 1 specialized center. Participants were randomized 1:1 to the intervention or control group. The intervention group was administered standard care and access to the DiaCert app at baseline and 3 months onward. The control group received standard care only. Outcomes of objectively measured physical activity using accelerometers, BMI, waist circumference, biomarkers, and blood pressure were assessed at baseline and follow-ups. Linear mixed models were used to assess differences in outcomes between the groups.

Results: A total of 181 study participants, 65.7% (119/181) men and 34.3% (62/181) women, were recruited into the study and randomized to the intervention (n=93) or control group (n=88). The participants’ mean age and BMI were 60.0 (SD 11.4) years and 30.4 (SD 5.3) kg/m², respectively. We found no significant effect of the intervention (group by time interaction) on MVPA at either the 3-month (β=1.51, 95% CI –5.53 to 8.55) or the 6-month (β=–3.53, 95% CI –10.97 to 3.92) follow-up. We found no effect on any of the secondary outcomes at follow-ups, except for a significant effect on BMI at 6 months (β=0.52, 95% CI 0.20 to 0.84). However, mean BMI did not differ between the groups at the 6-month follow-up.

Conclusions: We found no evidence that persons with type 2 diabetes being randomized to use an app promoting daily walking increased their levels of MVPA at 3 or 6 months’ follow-up compared with controls receiving standard care. The effect of the
app on BMI was unclear, and we found nothing to support an effect on secondary outcomes. Further research is needed to determine what type of mHealth intervention could be effective to increase physical activity among persons with type 2 diabetes.

**Trial Registration:** ClinicalTrials.gov NCT03053336; https://clinicaltrials.gov/study/NCT03053336


**KEYWORDS**

behavior change; exercise; intervention; mHealth; smartphone app; self-monitoring

**Introduction**

Despite increased knowledge and public health initiatives, more than 460 million people, corresponding to over 6% of the world’s population, are estimated to be diagnosed with type 2 diabetes today [1]. Persons with type 2 diabetes followed in primary care can be prescribed lifestyle interventions in combination with medications. Lifestyle interventions may include weight management, smoking cessation, stress reduction, and improved dietary habits or physical activity [2]. Physical activity is well known to have beneficial effects on glycemic control and to reduce risk factors for cardiovascular disease [3]. Yet, it has proven difficult to implement lifestyle interventions targeting physical activity in primary care [4]. Nevertheless, walking has been put forward as a useful therapeutic tool shown to improve glucose control, with clinically beneficial effects on blood glucose levels over time, and the potential to improve other clinical variables such as BMI and blood pressure in persons with type 2 diabetes [5].

During the past few decades, various telemonitoring, eHealth, and mobile health (mHealth) solutions targeting physical activity have been developed. Such solutions offer adaptable platforms for the delivery of self-management interventions that are easily accessible to both patients and health care practitioners, and users can engage with health information technology at their convenience. Smartphone apps may be useful in a health care setting to provide an additional tool to increase patients’ engagement through the use of self-monitoring of, for example, physical activity between routine visits [6].

Today, there are many commercially developed and available smartphone apps targeting self-management of chronic conditions. Common features of available apps targeted toward persons with type 2 diabetes include self-tracking of blood glucose levels and components targeting physical activity or diet in different ways [7,8]. Nevertheless, there is a wide variety in type and number of features for diabetes management in available apps [7], making it difficult for patients to select the most appropriate one to use. There are many commercial apps targeting lifestyle among persons with type 2 diabetes; however, few solutions primarily target physical activity within this group, and even fewer have been developed specifically for implementation in primary care. Therefore, we developed a digital platform and a smartphone app specifically for targeting physical activity among persons with type 2 diabetes treated within primary care [9]. The app was built to be integrated into the existing digital infrastructure of primary care in Sweden, with the aim to provide care givers and patients with a scientifically evaluated self-care management tool.

Results from studies evaluating the effectiveness of mHealth solutions, including smartphone apps, targeting persons with type 2 diabetes, have been summarized previously [10-15]. Significant reductions in hemoglobin A1c (HbA1c) levels are generally shown after 3 months of follow-up. Most of the evaluated apps allowed the user to monitor their blood glucose levels and included physical activity or diet, either alone or in combination, as additional features. mHealth interventions targeting physical activity in adults in the general population have been shown to increase both minutes of physical activity and steps per day [16-18]. Nevertheless, apps primarily targeting physical activity, without including a component of glucose monitoring, in persons with type 2 diabetes are relatively uncommon. Poppe et al [19] evaluated a self-regulation–based eHealth and mHealth intervention primarily targeting physical activity in persons with type 2 diabetes and found positive results of the intervention on increased activity and decreased sedentary behavior, whereas Thorsen et al [20] found no effect of app-based interval walking on MVPA over 52 weeks compared with standard care. In summary, there is still a need to develop interventions targeting physical activity that are effective and can be implemented into primary care.

The aim of this study was to evaluate the effects of the DiaCert smartphone app promoting daily walking on moderate to vigorous physical activity (MVPA) and clinical variables in persons with type 2 diabetes. Our primary aim was to test the hypothesis that the app would lead to an increase in minutes of MVPA at 3 months compared with standard care only. Our secondary aim was to assess the effect of the app on MVPA after 6 months and on the clinical variables BMI, waist circumference, HbA1c, cholesterol (total, low-density lipoprotein [LDL], and high-density lipoprotein [HDL]), triglycerides, and systolic and diastolic blood pressure at 3 and 6 months. We hypothesized that the app would lead to improvements in both MVPA and clinical variables.

**Methods**

**Trial Design**

We conducted a randomized controlled trial with 2 parallel groups between February 2017 and June 2019. The DiaCert study design [21] has been described in detail previously. Study participants were randomized 1:1 to the intervention or control group at baseline. The primary study outcome was MVPA (minutes/day) at 3 months of follow-up. Secondary outcomes included MVPA at 6 months of follow-up and the clinical variables BMI, waist circumference, HbA1c, cholesterol (total, LDL, and HDL), triglycerides, and systolic and diastolic blood pressure at 3 and 6 months of follow-up. No changes to methods...
were done after trial commencement. The study is reported according to the CONSORT (Consolidated Standards of Reporting Trials) statement [22] and the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) checklist, which is developed for eHealth or mHealth interventions [23].

Ethical Considerations
The trial was approved by the ethics committee of the regional ethical review board in Stockholm, Sweden (Dnr: 2016/2041-31/2, 2016/99-32, 2017/1406-32) and registered at ClinicalTrials.gov (NCT03053336). All participants received both oral and written information about the study and provided their written informed consent to participate. Participants received no compensation for participation in the study. After data collection, data were anonymized.

Study Participants
The inclusion criteria were (1) having a diagnosis of type 2 diabetes, (2) being 18 years of age or older, (3) being able to communicate in Swedish, and (4) having access to and being able to use a smartphone. The exclusion criterion was not being able to walk.

Patients were recruited continuously from 5 primary health care centers and 1 specialized medical center in the Stockholm area, Sweden. Patients at the participating centers received initial information about the study from their physician or diabetes nurse, and those interested in participating were contacted by study personnel and given more detailed information. Thereafter, patients either agreed to participate and were scheduled for a baseline introductory meeting, or declined participation. We did not record the number of patients who did not agree to participate. All study participants met with study personnel at baseline and after 3 and 6 months. On each occasion, study outcomes including physical activity and other lifestyle factors, as well as clinical variables, were assessed.

Interventions
Study participants randomized to the intervention group continued to receive standard care as usual but also downloaded the DiaCert app during the baseline meeting. To access the individual user account in the app, a personal 6-digit code was entered. The code was given to participants by study personnel 7 days after the baseline meeting to avoid overlap with baseline accelerometer measurements. The intervention group was encouraged to use the app daily for 3 months. At the 3-month follow-up meeting, participants deleted the app from their phones together with study personnel. Participants then received no intervention during 3 months and were offered access to the app again at the 6-month follow-up.

The DiaCert app displayed daily steps that through connection to a digital platform were shared with study personnel. An individual step goal between 1000 and 10,000 steps [24] was set at baseline based on the participant’s usual activity level. Participants in the intervention group were contacted by study personnel every second week by phone. During these follow-ups, the participant could revise his or her step goal with an even 500 steps increase or decrease. The maximum goal set at any time point was 10,000 steps. Users received automatic positive feedback messages including the user’s name in the app on days when the goal was met. In addition to daily steps, information on HbA1c taken during the study period was also displayed in the app.

The app design has previously been described in detail, including a presentation of the app screen by screen [9]. In brief, features of the app included a home screen displaying the daily steps in relation to an individual step goal during the past week. A circle was gradually filled as the user walked toward the step goal. It was completely filled and marked with a checkmark when the goal was reached. Through the home screen, the user was also able to access information on previous daily steps, questionnaires, and results of HbA1c. The app was continually updated to run with the current iOS and Android versions, but no changes were made to the content during the study. Users were asked to contact study personnel if they experienced malfunction of the app. The app was developed within the research project and is no longer available.

Study participants randomized to the control group received standard care, that is, continued their usual care as prescribed by their regular primary care physician and diabetes nurse, also after inclusion to the study. For ethical reasons, they were offered access to the DiaCert app at the 6-month follow-up.

Outcomes
Physical Activity
To assess the primary outcome of MVPA in this trial, physical activity was measured using the ActiGraph wGT3x-BT triaxial accelerometer (dynamic range: 8g) during 7 consecutive days. At each study meeting, the participants were asked to wear the accelerometer on their nondominant wrist day and night, starting at 4 PM the same day until 8 AM just over a week later. Participants wore the accelerometers on their wrist to increase feasibility and maximize compliance. Data were sampled at a frequency of 80 Hz.

We downloaded the collected data from each accelerometer using the manufacturer’s program (ActiLife Software, version 6.13.3; ActiGraph), and thereafter, the raw data were extracted for data processing. As suggested by Migueles et al [25], processing of accelerometer data was performed using the open-source R package GGIR. GGIR version 2.0-0, R version 3.6.1, and RStudio version 1.2.5019 were used. Data collected before the first and after the last midnight were excluded in order to examine 7 complete days. As the first step of analysis, data were averaged over 5-second epochs and aggregated through application of Euclidean norm minus 1, with negative values rounded up to 0. Autocalibration was performed using local gravity to adjust for calibration errors and unreliable signals.

The default cut point for MVPA (100 milligravity) and default settings for the definition and management of nonwear time (ie, 4x15 minutes) in GGIR were applied [26]. Nonwear time was by default replaced with imputed averaged activity from the same time the other measured days. A valid day was defined as at least 16 hours of wear time and a valid week required 4 valid
days (including at least 1 weekend day) [25]. Variables were weighted 5:2 with data collected on weekdays and weekend days. For MVPA, bouts of consistent activity lasting for at least 1 minute were used, where 80% of the included epochs had to be above or equal to the cut point [27].

In addition to accelerometer measurements, we also assessed daily physical activity at baseline with 2 validated general questions used in routine health care [28]. Participants were asked to (1) report their usual time spent exercising during a week and (2) add up and report the total time estimated spent doing other types of leisure time physical activities of lower intensity in bouts of at least 10 minutes during a week. Walking, cycling, or gardening was presented to the respondent to exemplify activity level.

Clinical Variables
A detailed description of measurements has been published elsewhere [21]. In brief, HbA1c (mmol/mol), total cholesterol (mmol/L), LDL cholesterol (mmol/L), HDL cholesterol (mmol/L), and triglycerides (mmol/L) were measured in fasting blood samples. HbA1c was measured using the reference measurement procedure by the International Federation for Clinical Chemistry and Laboratory Medicine [29]. The enzymatic method [30] was used to measure total cholesterol and HDL cholesterol. LDL cholesterol was calculated using the Friedewald equation [31].

Height (to the nearest 0.5 cm), body weight (to the nearest 0.1 kg), and waist circumference (to the nearest 1.0 cm) were measured by study personnel who also performed 1 manual assessment of blood pressure (systolic and diastolic) after the participant had been sitting down for at least 5 minutes. BMI was calculated based on measured height and weight (kg/m²).

Sample Size
Power calculations were performed a priori to determine the sample size need to detect a clinically significant difference of 8 minutes/day of MVPA [21]. A total of 250 participants (125 in each group) were estimated to provide 80% power at a 5% significance level. This included an expected dropout rate of 20%. Baseline data collection ended in June 2018 before reaching 250 participants because the DiaCert app was no longer compatible with the upgrades of iOS and Android.

Randomization and Blinding
A random allocation sequence list was generated by the first author (SEB) in Stata (version 14.0; StataCorp). Women and men were randomized separately in blocks of 10 within each participating primary health care center and the specialized medical center. Patients who agreed to participate in the study were randomized separately in blocks of 10 within each group allocation. Participants were blind to the intervention, neither participants nor study personnel were blinded to participants’ group allocation.

Statistical Methods
Baseline characteristics of study participants are presented as mean (SD) or number (%) for continuous and categorical variables, respectively. Variables were checked for normality and outliers. The Student t test or the chi-square test was used to assess potential differences in baseline characteristics between the intervention and control group.

We used linear mixed models with fixed and random intercept and slope for the time variables to assess if there were longitudinal differences in MVPA at 3 months of follow-up (primary outcome) and secondary outcomes including MVPA at 6 months of follow-up and BMI, waist circumference, HbA1c, total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, and systolic and diastolic blood pressure at 3 and 6 months of follow-up between the intervention and control group. We have therefore included, in addition to time and group terms, a group×time interaction term to assess if any differences in the study outcomes were constant at 3 and 6 months. For outcomes that differed significantly between the study groups at baseline, that is, MVPA, additional sensitivity analysis adjusting the models for baseline levels of the outcome was performed according to the methods described by Twisk et al [32]. The analysis of intervention effect was made following the intention-to-treat approach [33]. Missing data were associated with a primary care center, with 17/25 (68%) participants with missing accelerometer data at baseline belonging to primary care center 1. Missing data were assumed to be missing at random and not depending on the study group as the degree of missing data was similar in both the intervention and control group during the intervention. Participants with complete data at baseline for each specific outcome variable were included in the analysis of intervention effect at 3 and 6 months.

Post hoc sensitivity analyses using self-reported data on physical activity from the baseline questionnaire were carried out to deal with the unblinded nature of our study and the potential bias that may have been present during baseline accelerometer measurements. Although participants were not connected to the DiaCert app until after the completion of baseline accelerometer measurements, they were aware of which group they were randomized to during measurements. This could potentially have affected baseline levels of MVPA, which might not have represented the “true” levels of MVPA before the start of the study. Multiple imputation based on MICE (multiple imputation by chained equations) was used to address this issue [34]. We first set all baseline values of MVPA for the intervention group to missing. To predict MVPA at baseline for participants in the intervention group in the hypothetical scenario in which participants were blind to the intervention, we implemented MICE using all relevant baseline variables. The variables included in the model were age, sex, height, weight, education, household income, marital status, smoking and snuff habits, the year of diabetes diagnosis, the center of recruitment, self-reported levels of physical activity, and the number of valid accelerometer days at baseline. We compared the imputed and the observed values of MVPA at baseline using a 2-tailed t test. Linear mixed models, as described above, were thereafter fitted.
to contrast differences in MVPA between the intervention and control group using the imputed data for MVPA at baseline.

Statistical tests were 2-sided, and the significance level was set to $P<.05$. Statistical analyses were performed using Stata (version 17; StataCorp).

**Results**

A total of 181 persons with type 2 diabetes were included in the trial, of whom 93 were randomized to the intervention group and 88 were randomized to the control group. At 3 months, the dropout rate was 10.5% (19/181), and at 6 months it was 14.9% (27/181). Dropout was higher among participants in the intervention group (n=12 vs n=7 at 3 months, and n=6 vs n=2 at 6 months) than among those in the control group. In total, 156 participants had valid accelerometer data on physical activity at baseline. Of these, 137 (87.8%) also had valid data at the 3-month follow-up (primary outcome). At baseline, most participants (75.6%, 118/156) had valid accelerometer data from all 7 days, 14.7% (23/156) had valid data from 6 days, and 9.6% (15/156) from 5 or 4 days. A flowchart of participants with complete data from the different assessments is presented in Figure 1.

Figure 1. Flowchart of participation and data completeness at baseline and follow-up after 3 and 6 months in the DiaCert study. *Number of participants with valid accelerometer data, that is, at least 16 hours of wear time and a total of 4 valid days including at least 1 weekend day. †Triglycerides and cholesterol (total, low-density lipoprotein, and high-density lipoprotein). BP: blood pressure; HbA$_1c$: hemoglobin A$_1c$; WC: waist circumference.

Characteristics of all 181 study participants are presented in Table 1. The majority of participants were men (119/181, 65.8%), and the mean (SD) age at baseline among all participants was 60.0 (11.4) years. The mean (SD) BMI among all participants was 30.4 (5.3) kg/m$^2$. Overweight and obesity was common, 85.6% (155/181) had a BMI of $\geq 25$ kg/m$^2$, and 48.1% (87/181) had a BMI of $\geq 30$ kg/m$^2$. Mean (SD) HbA$_1c$ was 53.6 (12.8) mmol/mol. There were no statistically significant differences between the study groups regarding age distribution, primary care center belonging, sex, smoking, time since diabetes diagnosis, or clinical variables including HbA$_1c$ and lipid levels. In study participants with complete accelerometer data at baseline (n=156), there was a statistically significant difference in accelerometer-measured baseline levels of physical activity between the intervention and control group, with higher MVPA (38.3 vs 29.8 minutes/day, $P=.04$) measured in the intervention group. Using the imputed data for the intervention group, the baseline MVPA was estimated to be 32.3 minutes/day, which did not differ from the measured level in the control group ($P=.62$). Additionally, self-reported levels of physical activity at baseline, that is, time spent exercising and total leisure time activity, did not differ between the groups ($P=.20$ and $P=.20$, respectively). Baseline characteristics of participants with complete accelerometer data at baseline can be found in Multimedia Appendix 1.
Table 1. Characteristics of study participants by study group (N=181).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Intervention group (n=93)</th>
<th>Control group (n=88)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVPA&lt;sup&gt;b,c&lt;/sup&gt; (minutes/day)</td>
<td>76 (82) 38.3 (28.3)</td>
<td>80 (91) 29.7 (24.1)</td>
<td>.04</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>93 (100) 30.2 (5.5)</td>
<td>88 (100) 30.6 (5.2)</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Waist circumference</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>33 (35) 102 (12.6)</td>
<td>28 (32) 103 (15.8)</td>
<td>.75</td>
</tr>
<tr>
<td>Men</td>
<td>59 (63) 111 (15.6)</td>
<td>60 (68) 110 (13.3)</td>
<td>.75</td>
</tr>
<tr>
<td>Hemoglobin A&lt;sub&gt;1c&lt;/sub&gt; (mmol/mol)</td>
<td>89 (96) 53.6 (13.0)</td>
<td>88 (100) 53.5 (12.7)</td>
<td>.97</td>
</tr>
<tr>
<td>Total cholesterol (mmol/L)</td>
<td>74 (80) 4.56 (1.00)</td>
<td>70 (80) 4.50 (1.18)</td>
<td>.77</td>
</tr>
<tr>
<td>LDL&lt;sup&gt;d&lt;/sup&gt; cholesterol (mmol/L)</td>
<td>73 (78) 2.70 (0.89)</td>
<td>70 (80) 2.64 (1.13)</td>
<td>.72</td>
</tr>
<tr>
<td>HDL&lt;sup&gt;e&lt;/sup&gt; cholesterol (mmol/L)</td>
<td>74 (80) 1.25 (0.34)</td>
<td>70 (80) 1.23 (0.40)</td>
<td>.73</td>
</tr>
<tr>
<td>Triglycerides (mmol/L)</td>
<td>72 (77) 1.54 (0.76)</td>
<td>69 (78) 1.64 (0.94)</td>
<td>.45</td>
</tr>
<tr>
<td><strong>Blood pressure (mm Hg)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic</td>
<td>92 (99) 138.7 (16.2)</td>
<td>88 (100) 137.0 (14.8)</td>
<td>.45</td>
</tr>
<tr>
<td>Diastolic</td>
<td>92 (99) 83.6 (9.9)</td>
<td>88 (100) 82.6 (9.5)</td>
<td>.50</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td>Male</td>
<td>34 (37) —&lt;sup&gt;f&lt;/sup&gt;</td>
<td>28 (32) —</td>
<td>—</td>
</tr>
<tr>
<td>Female</td>
<td>59 (63) —</td>
<td>60 (68) —</td>
<td>—</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>&lt;50</td>
<td>20 (22) —</td>
<td>16 (18) —</td>
<td>—</td>
</tr>
<tr>
<td>50-59</td>
<td>28 (30) —</td>
<td>19 (22) —</td>
<td>—</td>
</tr>
<tr>
<td>60-69</td>
<td>31 (33) —</td>
<td>30 (34) —</td>
<td>—</td>
</tr>
<tr>
<td>≥70</td>
<td>14 (15) —</td>
<td>23 (26) —</td>
<td>—</td>
</tr>
<tr>
<td><strong>Leisure time activity&lt;sup&gt;g&lt;/sup&gt; (minutes/week)</strong></td>
<td></td>
<td></td>
<td>.27</td>
</tr>
<tr>
<td>&lt;60</td>
<td>6 (7) —</td>
<td>11 (13) —</td>
<td>—</td>
</tr>
<tr>
<td>60-90</td>
<td>8 (10) —</td>
<td>8 (10) —</td>
<td>—</td>
</tr>
<tr>
<td>90-150</td>
<td>17 (20) —</td>
<td>15 (18) —</td>
<td>—</td>
</tr>
<tr>
<td>150-300</td>
<td>14 (17) —</td>
<td>22 (26) —</td>
<td>—</td>
</tr>
<tr>
<td>&gt;300</td>
<td>39 (46) —</td>
<td>28 (33) —</td>
<td>—</td>
</tr>
<tr>
<td><strong>Primary care centers</strong></td>
<td></td>
<td></td>
<td>.88</td>
</tr>
<tr>
<td>1</td>
<td>35 (38) —</td>
<td>31 (35) —</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>14 (15) —</td>
<td>13 (15) —</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>6 (6) —</td>
<td>8 (9) —</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>23 (25) —</td>
<td>23 (26) —</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>10 (11) —</td>
<td>11 (13) —</td>
<td>—</td>
</tr>
<tr>
<td>Specialized medical centers</td>
<td>5 (5) —</td>
<td>2 (2) —</td>
<td>—</td>
</tr>
<tr>
<td><strong>Time spent exercising&lt;sup&gt;h&lt;/sup&gt; (minutes/week)</strong></td>
<td></td>
<td></td>
<td>.26</td>
</tr>
<tr>
<td>Never</td>
<td>35 (42) —</td>
<td>38 (44) —</td>
<td>—</td>
</tr>
<tr>
<td>&lt;30</td>
<td>12 (14) —</td>
<td>12 (14) —</td>
<td>—</td>
</tr>
<tr>
<td>30-90</td>
<td>13 (15) —</td>
<td>21 (24) —</td>
<td>—</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Intervention group (n=93)</td>
<td>Control group (n=88)</td>
<td>P value&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------------------------</td>
<td>----------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td></td>
<td>n (%)</td>
<td>Mean (SD)</td>
<td>n (%)</td>
</tr>
<tr>
<td>&gt;90</td>
<td>24 (29)</td>
<td>—</td>
<td>15 (17)</td>
</tr>
<tr>
<td>Smoking&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>11 (13)</td>
<td>—</td>
<td>9 (11)</td>
</tr>
<tr>
<td>No, ever smoker</td>
<td>31 (37)</td>
<td>—</td>
<td>40 (47)</td>
</tr>
<tr>
<td>No, never smoker</td>
<td>42 (50)</td>
<td>—</td>
<td>36 (42)</td>
</tr>
<tr>
<td>Time since diabetes diagnosis&lt;sup&gt;j&lt;/sup&gt; (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>8 (11)</td>
<td>—</td>
<td>13 (18)</td>
</tr>
<tr>
<td>1-5</td>
<td>20 (29)</td>
<td>—</td>
<td>17 (24)</td>
</tr>
<tr>
<td>&gt;5</td>
<td>42 (60)</td>
<td>—</td>
<td>42 (58)</td>
</tr>
<tr>
<td>Education&lt;sup&gt;k&lt;/sup&gt; (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤12</td>
<td>45 (54)</td>
<td>—</td>
<td>44 (51)</td>
</tr>
<tr>
<td>&gt;12</td>
<td>38 (46)</td>
<td>—</td>
<td>42 (49)</td>
</tr>
</tbody>
</table>

<sup>a</sup> 2-tailed t test was used for continuous variables and the chi-square test was used for categorical variables. Italicized P values represent statistical significance.

<sup>b</sup>MVPA: moderate to vigorous physical activity.

<sup>c</sup>Missing data from n=17 (intervention) and n=8 (control).

<sup>d</sup>LDL: low-density lipoprotein.

<sup>e</sup>HDL: high-density lipoprotein.

<sup>f</sup>—not available.

<sup>g</sup>From questionnaire, missing data n=9 (intervention) and n=4 (control).

<sup>h</sup>From questionnaire, missing data n=9 (intervention) and n=2 (control).

<sup>i</sup>Missing data n=9 (intervention) and n=3 (control).

<sup>j</sup>Missing data n=23 (intervention) and n=16 (control).

<sup>k</sup>Missing data n=10 (intervention) and n=2 (control).

**Effectiveness of the Intervention—MVPA**

Results from between-group analysis and the intervention effect on minutes/day of MVPA are shown in Table 2. The mean change in minutes/day of MVPA from baseline to follow-ups is graphically shown in Figure 2. The statistically significant difference in minutes/day of MVPA seen between the groups at baseline, with participants in the intervention group being more active than participants in the control group, remained at the 3-month follow-up. The predicted mean difference between the groups after 3 months was 10.05 minutes (95% CI 1.66-18.44). At the 6-month follow-up, there was no statistically significant difference in MVPA between the groups (β=5.02, 95% CI -3.72 to 13.75).
Table 2. The intervention effect on the primary outcome of daily minutes of moderate to vigorous physical activity (MVPA) at 3 months of follow-up and on secondary outcomes including MVPA at 6 months of follow-up and clinical variables in the DiaCert study.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Group sample means</th>
<th>Model estimates&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Group by time interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention (n=93)</td>
<td>Control (n=88)</td>
<td>Difference&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>n (%)</td>
<td>Mean (SD)</td>
<td>n (%)</td>
</tr>
<tr>
<td>MVPA (minutes/day)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 months</td>
<td>70 (75)</td>
<td>36.6 (25.5)</td>
<td>67 (76)</td>
</tr>
<tr>
<td>6 months</td>
<td>55 (59)</td>
<td>34.2 (29.4)</td>
<td>63 (72)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>93 (100)</td>
<td>30.2 (5.5)</td>
<td>88 (100)</td>
</tr>
<tr>
<td>3 months</td>
<td>81 (87)</td>
<td>30.1 (5.7)</td>
<td>79 (90)</td>
</tr>
<tr>
<td>6 months</td>
<td>70 (75)</td>
<td>30.0 (6.0)</td>
<td>79 (90)</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>92 (99)</td>
<td>107 (15.1)</td>
<td>88 (100)</td>
</tr>
<tr>
<td>3 months</td>
<td>80 (86)</td>
<td>107 (15.6)</td>
<td>79 (90)</td>
</tr>
<tr>
<td>6 months</td>
<td>69 (74)</td>
<td>107 (17.0)</td>
<td>79 (90)</td>
</tr>
<tr>
<td>Hemoglobin A&lt;sub&gt;1c&lt;/sub&gt; (mmol/mol)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>89 (96)</td>
<td>53.6 (13.0)</td>
<td>88 (100)</td>
</tr>
<tr>
<td>3 months</td>
<td>75 (81)</td>
<td>50.0 (9.9)</td>
<td>73 (83)</td>
</tr>
<tr>
<td>6 months</td>
<td>67 (72)</td>
<td>51.2 (10.7)</td>
<td>76 (86)</td>
</tr>
<tr>
<td>Total cholesterol (mmol/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>74 (80)</td>
<td>4.56 (1.00)</td>
<td>70 (80)</td>
</tr>
<tr>
<td>3 months</td>
<td>62 (67)</td>
<td>4.39 (0.80)</td>
<td>57 (65)</td>
</tr>
<tr>
<td>6 months</td>
<td>55 (59)</td>
<td>4.38 (0.84)</td>
<td>59 (67)</td>
</tr>
<tr>
<td>LDL&lt;sup&gt;d&lt;/sup&gt; cholesterol (mmol/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>73 (78)</td>
<td>2.70 (0.89)</td>
<td>70 (80)</td>
</tr>
<tr>
<td>3 months</td>
<td>59 (63)</td>
<td>2.44 (0.65)</td>
<td>56 (64)</td>
</tr>
<tr>
<td>6 months</td>
<td>54 (58)</td>
<td>2.44 (0.68)</td>
<td>57 (65)</td>
</tr>
<tr>
<td>HDL&lt;sup&gt;e&lt;/sup&gt; cholesterol (mmol/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>74 (80)</td>
<td>1.25 (0.34)</td>
<td>70 (80)</td>
</tr>
<tr>
<td>3 months</td>
<td>62 (67)</td>
<td>1.22 (0.34)</td>
<td>57 (65)</td>
</tr>
<tr>
<td>6 months</td>
<td>55 (59)</td>
<td>1.26 (0.36)</td>
<td>57 (65)</td>
</tr>
<tr>
<td>Triglycerides (mmol/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>72 (77)</td>
<td>1.54 (0.76)</td>
<td>69 (78)</td>
</tr>
<tr>
<td>3 months</td>
<td>60 (65)</td>
<td>1.71 (1.03)</td>
<td>58 (66)</td>
</tr>
<tr>
<td>6 months</td>
<td>52 (56)</td>
<td>1.56 (0.83)</td>
<td>56 (64)</td>
</tr>
<tr>
<td>Systolic BP&lt;sup&gt;f&lt;/sup&gt; (mm Hg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>92 (99)</td>
<td>139 (16.2)</td>
<td>88 (100)</td>
</tr>
<tr>
<td>3 months</td>
<td>80 (86)</td>
<td>135 (13.3)</td>
<td>79 (90)</td>
</tr>
<tr>
<td>6 months</td>
<td>70 (75)</td>
<td>136 (12.8)</td>
<td>78 (89)</td>
</tr>
<tr>
<td>Diastolic BP (mm Hg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>92 (99)</td>
<td>83.6 (9.9)</td>
<td>88 (100)</td>
</tr>
<tr>
<td>3 months</td>
<td>80 (86)</td>
<td>81.8 (9.1)</td>
<td>79 (90)</td>
</tr>
</tbody>
</table>
We found no statistically significant effect of the intervention (group by time interaction) on MVPA at either the 3- or the 6-month follow-up (Table 2). When adjusting for baseline levels of MVPA, results remained nonsignificant at both 3 (β=4.38, 95% CI –2.11 to 10.88) and 6 (β=–0.65, 95% CI –7.58 to 6.29) months. Additionally, results from sensitivity analyses using imputed baseline data also remained nonsignificant at follow-up after both 3 (β=6.86, 95% CI –4.05 to 17.78) and 6 (β=1.44, 95% CI –9.87 to 11.76) months.

**Effectiveness of Intervention—Clinical Variables**

Detailed results from between-group analyses and the intervention effect on clinical variables included as secondary outcomes are shown in Table 2. The mean change in outcomes from baseline to follow-ups is graphically shown in Multimedia Appendix 2. We found no statistically significant differences in any of the secondary outcomes at the 3- or 6-month follow-ups, except for in the analysis of BMI where a statistically significant effect of the intervention was seen at 6 months (group by time interaction: 0.52, 95% CI 0.20-0.84). However, there was no difference in mean BMI between the groups at the 6-month follow-up (predicted difference in mean: 0.11, 95% CI –1.45 to 1.67). Participants in the control group had a slightly higher BMI at baseline than participants in the intervention group (30.6 vs 30.2 kg/m\(^2\)), although this difference was not statistically significant (\(P=.61\)).

**Discussion**

**Principal Findings**

In this 2-armed randomized controlled trial, we found no clear effect of the DiaCert app promoting daily walking in persons with type 2 diabetes. We found no increase in objectively measured MVPA after neither 3 (primary outcome) nor 6 months of follow-up compared with standard care when accounting for baseline levels of MVPA. This is in line with results from Thorsen et al [20]. They were not able to show an effect of app-based interval walking on MVPA over 52 weeks compared with standard care among persons with type 2 diabetes. On the contrary, Poppe et al [19] found that an eHealth and mHealth intervention that primarily targeted physical activity in persons with type 2 diabetes led to increased physical activity and decreased sedentary behavior. Participants in our study and the study by Thorsen et al [20] reported more time in MVPA at baseline compared with those in the study by Poppe et al [19]. This might have contributed to the difference in effect between
investigated the effect of a basic self-monitoring app (tracking, feedback, information) and 2 theory-based tools (goal setting and point-based feedback) on MVPA among healthy but inactive adults. Four different groups received either a basic self-monitoring app only or the basic app together with (1) goal setting, or (2) feedback, or (3) goal setting and feedback. All groups increased their MVPA, but the feedback group showed the highest increase. While the DiaCert app comprised several behavioral change techniques that previously have been included in successful interventions, such as monitoring, goal setting, and feedback, it did not include other features associated with effective results, for example, frequent reminders or the option to share data with peers [40]. How the included components are designed may also affect results; for example, goal setting can be personalized or generic, and a goal can be more or less challenging.

**Strengths and Limitations**

The design of this randomized controlled study is one of the strengths of our study. The fact that study participants were recruited from 6 different care centers located in different areas with diverse populations and levels of socioeconomic status is another strength. This likely increased generalizability of our results. However, a limitation of our study is that we did not record the number of patients who turned down participation after being contacted by study personnel. Nevertheless, the mean age in our study was slightly lower than that of the average person with type 2 diabetes in Sweden, but levels of BMI and HbA\text{1c} were similar [41]. Younger persons may be more inclined to participate in an app-based intervention, although internet access and smartphone usage are high also among older age groups in Sweden [42]. Another strength of our study is that both men and women were included. Earlier studies have shown that women participate in physical activity programs more than men [43]. The larger proportion of men in our study could partly be explained by the higher prevalence of type 2 diabetes among men; almost 60% of persons with type 2 diabetes within primary care in Sweden are men [41]. It could also be speculated that older men are more interested in using technology than older women and therefore more likely to participate in an mHealth intervention.

The fact that participants were not blinded to the intervention is a limitation. While baseline information regarding the clinical variables is unlikely to have been affected by participants being aware of their group allocation, this knowledge may have had an impact on accelerometer measurements. Participants can, intentionally or unintentionally, change their physical activity behaviors during the measuring period. The statistically significant differences between the intervention and control group in accelerometer-measured MVPA at baseline were not seen for physical activity assessed in the baseline questionnaire. This could be interpreted as an immediate effect of being randomized to the intervention group and thereby feeling encouraged to become more active. Future studies should be careful not to disclose allocation information to participants until all baseline measures have been completed. It is also a limitation that personnel working in the study were not blinded during measurements. Nevertheless, objective accelerometer measurement of physical activity and clinical biomarkers.
including HbA1c and lipid levels represent a strength in our study design as they are less prone to biased estimates resulting from unblinded study personnel. The continuous recruitment of study participants during the whole year also reduced the risk of results being biased due to season.

We did not reach our goal of including 250 participants in the study, which is a limitation. Because a digital solution must be continuously updated to run with the current iOS and Android versions, we had to end recruitment after 2.5 years for practical reasons, before reaching the goal. However, dropout rates were lower than the estimated 20% in our power calculation [21]. Lack of data on user engagement and adherence to the app is a limitation as higher user engagement has also been associated with more favorable outcomes [44]. Decreased app engagement over time during an intervention period has been suggested as a potential explanation for the lack of more long-term effects of physical activity promoting apps [45].

Conclusions

We found no evidence that persons with type 2 diabetes being randomized to use an app promoting daily walking increased their levels of MVPA at 3 or 6 months of follow-up compared with controls receiving standard care. Further, the effect on BMI was unclear, and we found nothing to support an effect of being randomized to use the DiaCert app on waist circumference, HbA1c, total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, and systolic or diastolic blood pressure compared with standard care. Further research is needed to determine what type of mHealth physical activity intervention could be effective to increase physical activity and improve cardiometabolic markers among persons with type 2 diabetes.

Acknowledgments

We would like to express our gratitude to the personnel at the participating health care centers for the help in recruiting participants to our study. This work was supported by Stockholm County Council and Karolinska Institutet (4D; YTL), as well as through grants from the Swedish Research Council for Health, Working life, and Welfare (Dnr: 2016-00985; YTL). The study was also financed by grants provided by Region Stockholm (Dnr: 20170784, the NSV project and a clinical research appointment; YTL) and funding from the strategic research area in care sciences (SFO-V; SEB). The funders had no role in the design, analysis, or writing of this paper. Generative AI tools have not been used in any portion of the presented work or manuscript writing.

Data Availability

The data sets generated during and analyzed during this study are not publicly available due ethical restrictions but are available from the corresponding author on reasonable request.

Authors' Contributions

YTL is responsible for the DiaCert study and has designed the study together with SEB and ML. SEB, MH and CA have contributed significantly to the data collection. HE processed accelerometer data and SEB performed the statistical analysis together with GP with critical input from RB. SEB prepared the initial draft of the manuscript together with MH, which has been critically reviewed by all authors. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Characteristics of study participants with complete accelerometer data at baseline (n=156) by study group.

[DOCX File, 19 KB - ijmr_v13i1e53054_app1.docx ]

Multimedia Appendix 2

Changes over time in secondary outcomes in the intervention and control group. Predicted group means from linear mixed model analysis; results correspond to model estimates shown in Table 2.

[PNG File, 255 KB - ijmr_v13i1e53054_app2.png ]

Multimedia Appendix 3

CONSORT eHEALTH checklist.

[PDF File (Adobe PDF File), 50754 KB - ijmr_v13i1e53054_app3.pdf ]

References


Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth
HbA1c: hemoglobin A1c
HDL: high-density lipoprotein
LDL: low-density lipoprotein
mHealth: mobile health
MICE: multiple imputation by chained equations
MVPA: moderate to vigorous physical activity

©Stephanie E Bonn, Madeleine Hummel, Giulia Peveri, Helén Eke, Christina Alexandrou, Rino Bellocco, Marie Löf, Ylva Trolle Lagerros. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 21.03.2024. 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.
A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence–Based Models in Health Care Education and Practice: Development Study Involving a Literature Review

Malik Sallam¹,²,³, MD, PhD; Muna Barakat⁴, PhD; Mohammed Sallam⁵, MSc, PharmD

¹Department of Pathology, Microbiology and Forensic Medicine, School of Medicine, The University of Jordan, Amman, Jordan
²Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman, Jordan
³Department of Translational Medicine, Faculty of Medicine, Lund University, Malmo, Sweden
⁴Department of Clinical Pharmacy and Therapeutics, Faculty of Pharmacy, Applied Science Private University, Amman, Jordan
⁵Department of Pharmacy, Mediclinic Parkview Hospital, Mediclinic Middle East, Dubai, United Arab Emirates

Corresponding Author:
Malik Sallam, MD, PhD
Department of Pathology, Microbiology and Forensic Medicine
School of Medicine
The University of Jordan
Queen Rania Al-Abdullah Street-Aljubeiha
Amman, 11942
Jordan
Phone: 962 0791845186
Fax: 962 65353388
Email: malik.sallam@ju.edu.jo

Abstract

Background: Adherence to evidence-based practice is indispensable in health care. Recently, the utility of generative artificial intelligence (AI) models in health care has been evaluated extensively. However, the lack of consensus guidelines on the design and reporting of findings of these studies poses a challenge for the interpretation and synthesis of evidence.

Objective: This study aimed to develop a preliminary checklist to standardize the reporting of generative AI-based studies in health care education and practice.

Methods: A literature review was conducted in Scopus, PubMed, and Google Scholar. Published records with “ChatGPT,” “Bing,” or “Bard” in the title were retrieved. Careful examination of the methodologies employed in the included records was conducted to identify the common pertinent themes and the possible gaps in reporting. A panel discussion was held to establish a unified and thorough checklist for the reporting of AI studies in health care. The finalized checklist was used to evaluate the included records by 2 independent raters. Cohen κ was used as the method to evaluate the interrater reliability.

Results: The final data set that formed the basis for pertinent theme identification and analysis comprised a total of 34 records. The finalized checklist included 9 pertinent themes collectively referred to as METRICS (Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, and Specificity of prompts and language). Their details are as follows: (1) Model used and its exact settings; (2) Evaluation approach for the generated content; (3) Timing of testing the model; (4) Transparency of the data source; (5) Range of tested topics; (6) Randomization of selecting the queries; (7) Individual factors in selecting the queries and interrater reliability; (8) Count of queries executed to test the model; and (9) Specificity of the prompts and language used. The overall mean METRICS score was 3.0 (SD 0.58). The tested METRICS score was acceptable, with the range of Cohen κ of 0.558 to 0.962 (P < .001 for the 9 tested items). With classification per item, the highest average METRICS score was recorded for the “Model” item, followed by the “Specificity” item, while the lowest scores were recorded for the “Randomization” item (classified as suboptimal) and “Individual factors” item (classified as satisfactory).

Conclusions: The METRICS checklist can facilitate the design of studies guiding researchers toward best practices in reporting results. The findings highlight the need for standardized reporting algorithms for generative AI-based studies in health care, considering the variability observed in methodologies and reporting. The proposed METRICS checklist could be a preliminary tool for guiding the design and reporting of AI-based studies in health care.
Introduction

The integration of generative artificial intelligence (AI) models into health care education and practice holds promising perspectives with numerous possibilities for continuous improvement [1-5]. Examples of generative AI-based conversational models characterized by ease of use and perceived usefulness include ChatGPT by OpenAI, Bing by Microsoft, and Bard by Google [6-8].

The vast potential of generative AI-based models in health care can be illustrated as follows. First, generative AI-based models can facilitate streamlining of the clinical workflow, with subsequent improvement in efficiency manifested in reduced time for care delivery and reduced costs [1,9-11]. Second, generative AI-based models can enhance personalized medicine with a huge potential to achieve refined prediction of disease risks and outcomes [1,12,13]. Third, generative AI-based models can be implemented to improve health literacy among lay individuals through the provision of easily accessible and understandable health information [1,14,15].

Despite the aforementioned advantages of generative AI-based models in health care, several valid concerns were raised, which should be considered carefully owing to their serious consequences [1,4,16]. For example, the lack of clarity on how generative AI-based models are trained raises ethical concerns [17,18]. Additionally, these models have an inherent bias in the generated content based on the modality of training used for their development and updates [17,18]. Importantly, the generation of inaccurate or misleading content, which might appear scientifically plausible to nonexperts (referred to as “hallucinations”), could have profound negative impacts in health care settings [1,19-21]. Furthermore, the integration of generative AI-based models in health care could raise complex medicolegal and accountability questions, compounded by the issues of data privacy and cybersecurity risks [1,4,22,23].

Similarly, the use of generative AI-based models can cause a paradigm shift in information acquisition, particularly in health care education [1,24-26]. However, careful consideration of the best policies and practices to incorporate AI-based models in health care education is needed [27]. This issue involves the urgent need to address the issues of inaccuracies, possible academic dishonesty, decline in critical thinking development, and deterioration of practical training skills [1].

Recently, a remarkable number of studies investigated the applicability and disadvantages of prominent generative AI-based conversational models, such as ChatGPT, Microsoft Bing, and Google Bard, in various health care and educational settings [1,2,4,28-34]. However, synthesizing evidence from such studies can be challenging owing to several reasons. Variations in methodologies implemented in various studies as well as in the reporting standards is a major limitation. This issue could hinder the efforts aiming to compare and contrast the results of generative AI-based studies in health care, contributing to the complexity in this domain. This variability arises from several factors, including different settings of the tested models, prompt variability, varying approaches used to evaluate the generated content of generative AI-based models, varying range of tested topics, and possible bias in selecting the tested subjects. Additionally, variability can be related to the different number and varying expertise of individual raters of content quality, as well as the variable number of queries executed, among other factors [35-37].

Therefore, it is important to initiate and develop an approach that can aid in establishing standardized reporting practices for studies aiming to evaluate the content of generative AI-based models, particularly in health care. This standardization can be crucial to facilitate the design of generative AI-based studies in health care, ensuring rigor and achieving precise comparison and credible synthesis of findings across different studies. Thus, we aimed to propose a preliminary framework (checklist) to establish proper guidelines for the design and reporting of findings of generative AI-based studies that address health care–related topics.

Methods

Study Design

The study was based on a literature review to highlight the key methodological aspects in studies that investigated 3 generative AI-based models (ChatGPT, Bing, and Bard) in health care education and practice. The literature review was conducted to identify relevant literature indexed in databases up to November 11, 2023 [38]. The databases used for this literature search were Scopus, PubMed/MEDLINE, and Google Scholar.

Ethical Considerations

This study did not involve human subjects, and thus, the requirement of ethical permission was waived.

Literature Search to Identify Relevant Records

The Scopus string query was as follows: (TITLE-ABS-KEY (“artificial intelligence” OR “AI”) AND TITLE-ABS-KEY (“healthcare” OR “health care”) AND TITLE-ABS-KEY (“education” OR “practice”)) AND PUBYEAR > 2022 AND DOCSTYPE (ar OR re) AND (LIMIT-TO (PUBSTAGE , “final”)) AND (LIMIT-TO (SRCTYPE , “j”)) AND (LIMIT-TO (LANGUAGE , “English”)). The Scopus search yielded a total of 843 documents.
The PubMed advanced search tool was used as follows: (“artificial intelligence’’[Title/Abstract] OR “AI’’[Title/Abstract]) AND (“healthcare’’[Title/Abstract] OR “health care’’[Title/Abstract]) AND (“education’’[Title/Abstract] OR “practice’’[Title/Abstract]) AND (“2022/12/01’’[Date - Publication] : “2023/11/11’’[Date - Publication]). The PubMed search yielded a total of 564 records.

In Google Scholar and using the Publish or Perish software (version 8), in the title words and in the years 2022-2023, the search was as follows: “artificial intelligence” OR “AI” AND “healthcare” OR “health care” AND “education” OR “practice,” with a maximum of 999 records retrieved [39].

**Criteria for Record Inclusion**

The records from the 3 databases were merged using EndNote 20.2.1 software. This was followed by removal of duplicate records and removal of preprints by using the following function: ANY FIELD preprint OR ANY FIELD eprint OR ANY FIELD SSRN OR ANY FIELD Researchgate OR ANY FIELD researchsquare OR ANY FIELD. The retrieved records were eligible for the final screening step given the following inclusion criteria: (1) Original article; (2) English record; (3) Published (peer reviewed); and (4) Assessment in health care practice or health care education. Finally, the imported references were subjected to the search function in EndNote as follows: Title contains ChatGPT OR Title contains Bing OR Title contains Bard. The selection of the included records was performed by the first author (Malik Sallam).

**Development of the Initial Checklist Items**

Initial development of the proposed checklist began with the assessment of the methodology and results sections of the included records, a majority of which were regarded as cross-sectional descriptive studies. Then, we referred to 2 commonly used reporting and quality guidelines to proactively explore pertinent themes for the proposed checklist based on the nature of the included records: (1) STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) Statement: guidelines for reporting observational studies (checklist: cross-sectional studies) and (2) CASP (Critical Appraisal Skills Programme) Checklist (CASP qualitative studies checklist) [40,41]. This facilitated the allocation of ethical considerations, including transparency, methodological rigor, and issues related to bias, in the proposed checklist. Then, the 3 authors conducted an independent content review process to identify all the possible essential themes and best practices in generative AI-based health care studies among the included records. Finally, the authors had a collaborative discussion to refine the selected themes and classify these themes into specific components relevant to the study objectives. Special attention was given to aspects of method and result reporting that were perceived to impact the quality and reproducibility of the records, as identified by the 3 authors.

**Establishing the Final Checklist Criteria**

Careful examination of the included records resulted in the compilation of 3 independent lists of “pertinent themes,” which are herein defined as being critical or recurring in the reporting of results of generative AI-based studies. A thorough discussion among the authors followed to reach a consensus on the pertinent themes. Recurring themes were defined as those found in the methods of at least three separate records. Critical aspects were defined as those that would impact the conclusions of the included records as agreed by the 3 authors.

The final pertinent themes were selected based on their author-perceived significance in the quality and reproducibility of the findings. A final list of 9 themes was agreed upon by the authors as follows: (1) the “Model” of the generative AI-based tool or tools used in the included record and the explicit mention of the exact settings employed for each tool; (2) the “Evaluation” approach to assess the quality of the content generated by the generative AI-based model in terms of objectivity to reach unbiased findings and subjectivity; (3) the exact “Timing” of generative AI-based model testing and its duration; (4) the “Transparency” of data sources used to generate queries for the generative AI-based model testing, including the permission to use copyrighted content; (5) the “Range” of topics tested (single topic, multiple related topics, or various unrelated topics, as well as the breadth of intertopic and intratopic queries tested); (6) the degree of “Randomization” of topics selected to be tested to consider the potential bias; (7) the “Individual” subjective role in evaluating the content and the possibility of interrater reliability concordance or discordance; (8) the number (“Count”) of queries executed on each generative AI model entailing the sample size of queries tested; and (9) the “Specificity” of prompts used on each generative AI-based model, including the exact phrasing of each prompt and the presence of feedback and learning loops, and the “Specificity” of the language or languages used in testing, besides any other cultural issues. Thus, the final checklist was termed METRICS (Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, and Specificity of prompts and language).

**Scoring the METRICS Items and Classification of the METRICS Score**

Testing of the included records was performed by 2 independent raters (Malik Sallam and Mohammed Sallam) independently, with each METRICS item scored using a 5-point Likert scale as follows: 5=excellent, 4=very good, 3=good, 2=satisfactory, and 1=suboptimal. For the items that were deemed “not applicable” (eg, individual factors for studies that employed objective methods for evaluation), no score was given. The scores for the 2 raters were then averaged. The average METRICS score was calculated as the sum of average scores for each applicable item divided by 10 minus the number of items deemed not applicable.

The subjective assessment of the 2 raters was performed based on predefined criteria as a general guide. For example, if the exact dates of model queries were mentioned, the “Timing” item was scored as excellent. The count of queries was agreed to be categorized as excellent if it was more than 500, while a single case or no mention of the count was considered suboptimal. For the prompt attributes, scores were assigned based on the availability of exact prompts, explicit mention of the language used, and details of prompting. Thus, prompts and language specificity were appraised positively if the study clearly and explicitly made the exact prompts available and if
there was an explicit mention of the language employed in the prompts. The evaluation method was agreed to be rated higher for objective assessments with full details and lower for subjective assessments. The explicit mention of the method of interrater reliability testing was agreed to be scored higher for the “Individual” item. Transparency was assessed based on the comprehensiveness of the data source, and the presence of full database disclosure and permission to use the data was agreed to be given an excellent score. Randomization was agreed to be scored the lowest for the absence of details and the highest for explicit detailed descriptions.

Finally, we decided to highlight the records that scored the highest for each METRICS item. The decision to take this approach was based on an attempt to refrain from providing examples for the other quality categories to avoid premature conclusions regarding the quality of the included studies owing to the preliminary pilot nature of the METRICS tool.

Statistical and Data Analysis

The average METRICS scores were classified into distinct categories of equal weights as follows: excellent (score 4.21-5.00), very good (3.41-4.20), good (2.61-3.40), satisfactory (1.81-2.60), and suboptimal (1.00-1.80).

The Cohen κ measure was used to assess the interrater reliability by 2 independent raters. The Cohen κ measure was categorized as follows: <0.20, poor agreement; 0.21-0.40, fair agreement; 0.41-0.60, good agreement; 0.61-0.80, very good agreement; and 0.81-1.00, excellent agreement.

For statistical analysis, we used IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp). A P value <.05 was considered significant.

Results

Description of the Included Studies

A total of 34 studies were included in the final analysis that aimed to establish the METRICS criteria (Figure 1).

The most common source of records was Cureus, with 9 of the 34 records (27%), followed by BMJ Neurology Open, with 2 of the 34 records (6%). The remaining 23 records were published in 23 different journals.

Evaluation of the Included Records Based on MED-METRICS Items

The METRICS checklist was divided into the following 3 parts: “Model” attributes, “Evaluation” approach, and features of “Data” (MED-METRICS).

The complete details of the model attributes of the included studies are presented in Table 1.

ChatGPT was tested in all the records included (34/34, 100%), followed by Google Bard (5/34, 15%) and Bing Chat (5/34, 15%). The exact dates of generative AI-based model queries were explicitly mentioned in 13 of the 34 records (38%). The count of cases or questions that were tested in the studies ranged from a single case to 2576 questions. The majority of studies (23/34, 68%) tested the AI models based on queries in the English language.

The complete details of the evaluation approach of the content generated by the AI-based models in the included studies are presented in Table 2.

Objective evaluation of the content generated by the generative AI-based model was noted in 15 of the 34 records (44%).

The complete details of the features of data used to generate queries for testing on the generative AI-based models, including the range of topics and randomization, are presented in Table 3.

Explicit mention of the randomization process was only noted in 4 of the 34 included studies (9%). Of the 34 records, 6 (18%) involved broad multidisciplinary medical exam questions (18%). Moreover, 2 studies (6%) explicitly mentioned the permission to use the data for the studies.
Table 1. Details of the model attributes of the included studies.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Model</th>
<th>Timing</th>
<th>Count</th>
<th>Specificity of the prompts and language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Ashwal et al [42]</td>
<td>ChatGPT-3.5, ChatGPT-4, Bing, and Bard with unclear settings</td>
<td>One month, May 2023</td>
<td>255 drug-drug pairs</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Alfertshofer et al [43]</td>
<td>ChatGPT with unclear settings</td>
<td>Not provided</td>
<td>1800 questions</td>
<td>The exact prompt was used for each question. The questions were taken from United States, United Kingdom, Italian, French, Spanish, and Indian exams. A new session was used for each question.</td>
</tr>
<tr>
<td>Ali et al [44]</td>
<td>ChatGPT Feb 9 free version with unclear settings</td>
<td>Not provided</td>
<td>50 items</td>
<td>Information was provided fully in the supplementary file of the article in English.</td>
</tr>
<tr>
<td>Aljindan et al [45]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>220 questions</td>
<td>Initial prompting that involved role playing as a medical professional. The language was English.</td>
</tr>
<tr>
<td>Altamimi et al [46]</td>
<td>ChatGPT-3.5</td>
<td>Single day not otherwise specified</td>
<td>9 questions</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Baglivo et al [47]</td>
<td>Bing, ChatGPT, Chatsonic, Bard, and YouChat with full details of the mode and large language model, including plugins</td>
<td>Exact dates were provided for each model (April 12, 13, and 14, 2023, and July 13, 2023)</td>
<td>15 questions</td>
<td>Italian was used.</td>
</tr>
<tr>
<td>Biswas et al [48]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>Exact date provided (March 16, 2023)</td>
<td>11 questions</td>
<td>The exact prompts were provided in English, with a new session for each question.</td>
</tr>
<tr>
<td>Chen et al [49]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>560 questions</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Deiana et al [50]</td>
<td>ChatGPT-3.5 and ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>11 questions</td>
<td>The exact prompts were not explicitly provided. English was used. A new session was used for each question. Up to three iterations were allowed for incorrect responses.</td>
</tr>
<tr>
<td>Fuchs et al [51]</td>
<td>ChatGPT-3 and ChatGPT-4 with unclear settings</td>
<td>Exact dates provided (February 19, 2023, and March 25, 2023)</td>
<td>60 questions</td>
<td>Dental medicine questions were translated from German to English, while the other questions were already present in English. Exact prompts were provided in English with prompting in 2 groups for the same questions: one group was primed with instructions, while the other was not primed. A total of 20 trials were conducted per group, and chat history was cleared after each trial.</td>
</tr>
<tr>
<td>Ghosh &amp; Bir [52]</td>
<td>ChatGPT (version March 14, 2023) with unclear settings</td>
<td>March 14 and 16, 2023</td>
<td>200 questions</td>
<td>The exact prompts and language were not explicitly provided. The first response was taken as final, and the option of “regenerate response” was not used.</td>
</tr>
<tr>
<td>Giannos [53]</td>
<td>ChatGPT-3 and ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>69 questions</td>
<td>Not explicitly provided.</td>
</tr>
<tr>
<td>Gobira et al [54]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>125 questions</td>
<td>Portuguese was used.</td>
</tr>
<tr>
<td>Grewal et al [55]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>The first week of May 2023</td>
<td>Not clear</td>
<td>The exact prompts were provided in English. One follow-up prompt was used for enhancement of some prompts.</td>
</tr>
<tr>
<td>Guerra et al [56]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>591 questions</td>
<td>The exact prompts were provided, while the language was not explicitly provided.</td>
</tr>
<tr>
<td>Hamed et al [57]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>Not clear</td>
<td>The exact prompts and language were not explicitly provided. Different prompts were tried to identify the most suitable.</td>
</tr>
<tr>
<td>Hoch et al [58]</td>
<td>ChatGPT (May 3rd version) with unclear settings</td>
<td>May 5 and 7, 2023</td>
<td>2576 questions</td>
<td>The exact prompts were provided, while the language was not explicitly provided.</td>
</tr>
<tr>
<td>Authors</td>
<td>Model</td>
<td>Timing</td>
<td>Count</td>
<td>Specificity of the prompts and language</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Juhi et al [59]</td>
<td>ChatGPT with unclear settings</td>
<td>February 20, 2023, to March 5, 2023</td>
<td>40 drug-drug pairs</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Kuang et al [60]</td>
<td>ChatGPT with unclear settings</td>
<td>Not provided</td>
<td>Not clear</td>
<td>The exact prompts were not explicitly provided. English was used.</td>
</tr>
<tr>
<td>Kumari et al [61]</td>
<td>ChatGPT-3.5, Bard, and Bing with unclear settings</td>
<td>July 30, 2023</td>
<td>50 questions</td>
<td>The exact prompts were not explicitly provided. English was used.</td>
</tr>
<tr>
<td>Kung et al [62]</td>
<td>ChatGPT-3.5 and ChatGPT-4 with unclear settings</td>
<td>July 2023</td>
<td>215 questions</td>
<td>Not clear</td>
</tr>
<tr>
<td>Lai et al [63]</td>
<td>ChatGPT-4 (May 24; Version 3.5) with unclear settings</td>
<td>Not provided</td>
<td>200 questions</td>
<td>The exact prompts and language were not explicitly provided. Three attempts to answer the complete set of questions over 3 weeks (once per week), with a new session for each question.</td>
</tr>
<tr>
<td>Lyu et al [64]</td>
<td>ChatGPT with unclear settings</td>
<td>Mid-February 2023</td>
<td>Not clear</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Moise et al [65]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>Not provided</td>
<td>23 questions</td>
<td>The exact prompts were provided in English, with a new session for each question.</td>
</tr>
<tr>
<td>Oca et al [66]</td>
<td>ChatGPT, Bing, and Bard with unclear settings</td>
<td>April 11, 2023</td>
<td>20 queries for each model</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Oztermeli &amp; Oztermeli [67]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>Not provided</td>
<td>1177 questions</td>
<td>The exact prompts were not explicitly provided. Turkish was used, with a new session for each question.</td>
</tr>
<tr>
<td>Pugliese et al [68]</td>
<td>ChatGPT with unclear settings</td>
<td>March 25, 2023</td>
<td>15 questions</td>
<td>The exact prompts were provided in English, with a new session for each question.</td>
</tr>
<tr>
<td>Sallam et al [69]</td>
<td>ChatGPT (default model) with unclear settings</td>
<td>February 25, 2023</td>
<td>Not provided</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Seth et al [70]</td>
<td>ChatGPT-3.5, Bard, and Bing AI</td>
<td>Not provided</td>
<td>6 questions</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Suthar et al [71]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>140 cases</td>
<td>The exact prompts were not explicitly provided. English was used.</td>
</tr>
<tr>
<td>Walker et al [72]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>5 cases</td>
<td>The exact prompts were not explicitly provided. English was used, with a new session for each question.</td>
</tr>
<tr>
<td>Wang et al [73]</td>
<td>ChatGPT-3.5 and ChatGPT-4 with unclear settings</td>
<td>February 14, 2023, for ChatGPT-3.5 and May 14-16, 2023, for ChatGPT-4</td>
<td>300 questions</td>
<td>The exact prompts were provided. Chinese and English were used. The prompts were enhanced though role play.</td>
</tr>
<tr>
<td>Wang et al [74]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>March 5-10, 2023</td>
<td>Not clear</td>
<td>Chinese (Mandarin) and English were used. Examples of prompts were provided.</td>
</tr>
<tr>
<td>Zhou et al [75]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>April 24-25, 2023</td>
<td>Single case and multiple poll questions</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Authors</td>
<td>Evaluation of performance</td>
<td>Individual role and interrater reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-Ashwal et al [42]</td>
<td>Objective via 2 different clinical reference tools</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alfertshofer et al [43]</td>
<td>Objective based on the key answers, with the questions screened independently by 4 investigators</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ali et al [44]</td>
<td>Objective for multiple-choice questions and true or false questions, and subjective for short-answer and assay questions</td>
<td>Assessment by 2 assessors independently with intraclass correlation coefficient for agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aljindan et al [45]</td>
<td>Objective based on key answers and historical performance metrics</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altamimi et al [46]</td>
<td>Subjective</td>
<td>Not clear; Assessment for accuracy, informativeness, and accessibility by clinical toxicologists and emergency medicine physicians</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baglivo et al [47]</td>
<td>Objective based on key answers and comparison with 5th year medical students’ performance</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biswas et al [48]</td>
<td>Subjective by a 5-member team of optometry teaching and expert staff with over 100 years of clinical and academic experience between them; Independent evaluation on a 5-point Likert scale ranging from very poor to very good</td>
<td>The median scores across raters for each response were studied; The score represented rater consensus, while the score variance represented disagreements between the raters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen et al [49]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deiana et al [50]</td>
<td>Subjective based on qualitative assessment of correctness, clarity, and exhaustiveness; Each response rated using a 4-point Likert scale scored from strongly disagree to strongly agree</td>
<td>Independent assessment by 2 raters with experience in vaccination and health communication topics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuchs et al [51]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghosh &amp; Bir [52]</td>
<td>Objective based on key answers; Subjectivity by raters’ assessment</td>
<td>Scoring by 2 assessors on a scale of 0 to 5, with 0 being incorrect and 5 being fully correct, based on a preselected answer key</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giannos [53]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gobira et al [54]</td>
<td>Objective based on key answers, with an element of subjectivity through classifying the responses as adequate, inadequate, or indeterminate</td>
<td>Two raters independently scored the accuracy; After individual evaluations, the raters performed a third assessment to reach a consensus on the questions with differing results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grewal et al [55]</td>
<td>Not clear</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guerra et al [56]</td>
<td>Subjective through comparison with the results of a previous study on the average performance of users, and a cohort of medical students and neurosurgery residents</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hamed et al [57]</td>
<td>Subjective</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoch et al [58]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juhi et al [59]</td>
<td>Subjective and the use of Stockley’s Drug Interactions Pocket Companion 2015 as a reference key</td>
<td>Two raters reached a consensus for categorizing the output</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuang et al [60]</td>
<td>Subjective</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kumari et al [61]</td>
<td>Subjective; Content validity checked by 2 experts of curriculum design</td>
<td>Three independent raters scored content based on their correctness, with an accuracy score ranging from 1 to 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kung et al [62]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lai et al [63]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lyu et al [64]</td>
<td>Subjective</td>
<td>Two experienced radiologists (with 21 and 8 years of experience) evaluated the quality of the ChatGPT responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moise et al [65]</td>
<td>Subjective through comparison with the latest American Academy of Otolaryngology–Head and Neck Surgery Foundation Clinical Practice Guideline: Tymanostomy Tubes in Children (Update)</td>
<td>Two independent raters evaluated the output; The interrater reliability was assessed using the Cohen κ test; To confirm consensus, responses were reviewed by the senior author</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oca et al [66]</td>
<td>Not clear</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Evaluation of performance</td>
<td>Individual role and interrater reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------</td>
<td>------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oztermeli &amp; Oztermeli [67]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pugliese et al [68]</td>
<td>Subjective using the Likert scale for accuracy, completeness, and comprehensiveness</td>
<td>Multirater: 10 key opinion leaders in nonalcoholic fatty liver disease and 1 nonphysician with expertise in patient advocacy independently rating the AI(^a) content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sallam et al [69]</td>
<td>Subjective based on correctness, clarity, and conciseness</td>
<td>Fleiss multirater $\kappa$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seth et al [70]</td>
<td>Subjective through comparison with the current health care guidelines for rhinoplasty, and evaluation by a panel of plastic surgeons through a Likert scale to assess the readability and complexity of the text and the education level required for understanding, and the modified DISCERN(^b) score</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suthar et al [71]</td>
<td>Subjective by 3 fellowship-trained neuroradiologists, using a 5-point Likert scale, with 1 indicating “highly improbable” and 5 indicating “highly probable”</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walker et al [72]</td>
<td>Modified EQIP(^c) Tool with comparison with UK National Institute for Health and Care Excellence guidelines for gallstone disease, pancreatitis, liver cirrhosis, or portal hypertension, and the European Association for Study of the Liver guidelines</td>
<td>All answers were assessed by 2 authors independently, and in case of a contradictory result, resolution was achieved by consensus; The process was repeated 3 times per EQIP item; Wrong or out of context answers, known as “AI hallucinations,” were recorded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al [73]</td>
<td>Subjective</td>
<td>Unclear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al [74]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou et al [75]</td>
<td>Subjective</td>
<td>Unclear</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)AI: artificial intelligence.

\(^b\)DISCERN is an instrument for judging the quality of written consumer health information on treatment choices.

\(^c\)EQIP: Ensuring Quality Information for Patients.
Table 3. Classification of the included studies based on the features of data used to generate queries for testing on the generative artificial intelligence–based models.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Transparency</th>
<th>Range</th>
<th>Randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Ashwal et al [42]</td>
<td>Full description using 2 tools for assessment: Micromedex, a subscription-based drug-drug interaction screening tool, and Drugs.com, a free database</td>
<td>Narrow; Drug-drug interaction prediction</td>
<td>Nonrandom; Purposeful selection of the drugs by a clinical pharmacist; 5 drugs paired with the top 51 prescribed drugs</td>
</tr>
<tr>
<td>Alferthofer et al [43]</td>
<td>Full description using the question bank AMBOSS, with official permission for the use of the AMBOSS USMLE step 2CK practice question bank for research purposes granted by AMBOSS</td>
<td>Broad</td>
<td>Randomly extracted</td>
</tr>
<tr>
<td>Ali et al [44]</td>
<td>Developed by the researchers and reviewed by a panel of experienced academics for accuracy, clarity of language, relevance, and agreement on correct answers; Evaluation of face validity, accuracy, and suitability for undergraduate dental students</td>
<td>Narrow intersubject (dentistry); Broad intra-subject in restorative dentistry, periodontics, fixed prosthodontics, removable prosthodontics, endodontics, pedodontics, orthodontics, preventive dentistry, oral surgery, and oral medicine</td>
<td>Not clear</td>
</tr>
<tr>
<td>Aljindan et al [45]</td>
<td>Saudi Medical Licensing Exam questions extracted from the subscription CanadaQBank website</td>
<td>Broad in medicine, with 30% of the questions from medicine, 25% from obstetrics and gynecology, 25% from pediatrics, and the remaining 20% from surgery</td>
<td>Randomized through 4 researchers to ensure comprehensive coverage of questions and eliminate potential bias in question selection</td>
</tr>
<tr>
<td>Altamimi et al [46]</td>
<td>Snakebite management information guidelines derived from the World Health Organization, Centers for Disease Control and Prevention, and the clinical literature</td>
<td>Narrow</td>
<td>Not clear</td>
</tr>
<tr>
<td>Baglivo et al [47]</td>
<td>The Italian National Medical Residency test</td>
<td>Narrow; Vaccination-related questions from the Italian National Medical Residency Test</td>
<td>Not clear</td>
</tr>
<tr>
<td>Biswas et al [48]</td>
<td>Constructed based on the frequently asked questions on the myopia webpage of the Association of British Dispensing Opticians and the College of Optometrists</td>
<td>Narrow involving 9 categories: 1 each for disease summary, cause, symptom, onset, prevention, complication, natural history of untreated myopia, and prognosis, and 3 involving treatments</td>
<td>Not clear</td>
</tr>
<tr>
<td>Chen et al [49]</td>
<td>BoardVitals, which is an online question bank accredited by the Accreditation Council for Continuing Medical Education</td>
<td>Neurology-based; Broad intrasubject: basic neuroscience; behavioral, cognitive, psychiatry; cerebrovascular; child neurology; congenital; cranial nerves; critical care; de-myelinating disorders; epilepsy and seizures; ethics; genetic; headache; imaging or diagnostic studies; movement disorders; neuro-ophthalmology; neuro-otology; neuroinfectious disease; neurologic complications of systemic disease; neuromuscular; neurotoxicology, nutrition, metabolic; oncology; pain; pharmacology; pregnancy; sleep; and trauma</td>
<td>Not clear</td>
</tr>
<tr>
<td>Deiana et al [50]</td>
<td>Questions concerning vaccine myths and misconceptions by the World Health Organization</td>
<td>Narrow on vaccine myths and misconceptions</td>
<td>Not clear</td>
</tr>
<tr>
<td>Fuchs et al [51]</td>
<td>Digital platform self-assessment questions tailored for dental and medical students at the University of Bern’s Institute for Medical Education</td>
<td>Broad with multiple-choice questions designed for dental students preparing for the Swiss Federal Licensing Examination in Dental Medicine, and allergists and immunologists preparing for the European Examination in Allergy and Clinical Immunology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Ghosh &amp; Bir [52]</td>
<td>Department question bank, which is a compilation of first and second semester questions from various medical universities across India</td>
<td>Biochemistry</td>
<td>Random without details of randomization</td>
</tr>
<tr>
<td>Giannos [53]</td>
<td>Specialty Certificate Examination Neurology Web Questions bank</td>
<td>Neurology and neuroscience</td>
<td>Not clear</td>
</tr>
<tr>
<td>Authors</td>
<td>Transparency</td>
<td>Range</td>
<td>Randomization</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Gobira et al [54]</td>
<td>Not clear</td>
<td>National Brazilian Examination for Revalidation of Medical Diplomas issued by Foreign Higher Education Institutions (Revalida)</td>
<td>Not clear</td>
</tr>
<tr>
<td>Grewal et al [55]</td>
<td>Not clear</td>
<td>Preventive medicine, gynecology and obstetrics, surgery, internal medicine, and pediatrics</td>
<td>Not clear</td>
</tr>
<tr>
<td>Guerra et al [56]</td>
<td>Not clear</td>
<td>Neurosurgery across 7 subspecialties: tumor, cerebrovascular, trauma, spine, functional, pediatrics, and pain or nerve</td>
<td>Not clear</td>
</tr>
<tr>
<td>Hamed et al [57]</td>
<td>Not clear</td>
<td>Management of diabetic ketoacidosis</td>
<td>Not clear</td>
</tr>
<tr>
<td>Hoch et al [58]</td>
<td>Not clear</td>
<td>Otolaryngology with a range of 15 distinct otolaryngology subspecialties, including allergology, audiology, ear, nose and throat, tumors, face and neck, inner ear and skull base, larynx, middle ear, oral cavity and pharynx, nose and sinuses, phoniatrics, salivary glands, sleep medicine, vestibular system, and legal aspects</td>
<td>Not clear</td>
</tr>
<tr>
<td>Juhi et al [59]</td>
<td>Not clear</td>
<td>Narrow on drug-drug interaction</td>
<td>Not clear</td>
</tr>
<tr>
<td>Kuang et al [60]</td>
<td>Not clear</td>
<td>Neurosurgery</td>
<td>Not clear</td>
</tr>
<tr>
<td>Kumari et al [61]</td>
<td>Not clear</td>
<td>Hematology with the following intrasubject aspects: case solving, laboratory calculations, disease interpretations, and other relevant aspects of hematology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Kung et al [62]</td>
<td>Not clear</td>
<td>Orthopedics</td>
<td>Not clear</td>
</tr>
<tr>
<td>Lai et al [63]</td>
<td>Not clear</td>
<td>The United Kingdom Medical Licensing Assessment, which is a newly derived national undergraduate medical exit examination</td>
<td>Not clear</td>
</tr>
<tr>
<td>Lyu et al [64]</td>
<td>Not clear</td>
<td>Broad in medicine with the following aspects: acute and emergency, cancer, cardiovascular, child health, clinical hematology, ear, nose and throat, endocrine and metabolic, gastrointestinal including liver, general practice and primary health care, genetics and genomics, infection, medical ethics and law, medicine of older adults, mental health, musculoskeletal, neuroscience obstetrics and gynecology, ophthalmology, palliative and end of life care, perioperative medicine and anesthesia, renal and urology, respiratory, sexual health, social and population health, and surgery</td>
<td>Not clear</td>
</tr>
<tr>
<td>Moise et al [65]</td>
<td>Not clear</td>
<td>Narrow involving otolaryngology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Oca et al [66]</td>
<td>Not clear</td>
<td>Narrow involving only queries on accurate recommendation of close ophthalmologists</td>
<td>Not clear</td>
</tr>
<tr>
<td>Authors</td>
<td>Transparency</td>
<td>Range</td>
<td>Randomization</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>-------</td>
<td>---------------</td>
</tr>
<tr>
<td>Oztermeli &amp; Oztermeli [67]</td>
<td>Turkish medical specialty exam, prepared by the Student Selection and Placement Center</td>
<td>Broad: basic sciences including anatomy, physiology-histology-embryology, biochemistry, microbiology, pathology, and pharmacology; clinical including internal medicine, pediatrics, general surgery, obstetrics and gynecology, neurology, neurosurgery, psychiatry, public health, dermatology, radiology, nuclear medicine, otolaryngology, ophthalmology, orthopedics, physical medicine and rehabilitation, urology, pediatric surgery, cardiovascular surgery, thoracic surgery, plastic surgery, anesthesiology and reanimation, and emergency medicine</td>
<td>Not clear</td>
</tr>
<tr>
<td>Pugliese et al [68]</td>
<td>Expert selection of 15 questions commonly asked by patients with nonalcoholic fatty liver disease</td>
<td>Narrow involving nonalcoholic fatty liver disease aspects</td>
<td>Not clear</td>
</tr>
<tr>
<td>Sallam et al [69]</td>
<td>Panel discussion of experts in health care education</td>
<td>Broad on health care education, medical, dental, pharmacy, and public health</td>
<td>Not clear</td>
</tr>
<tr>
<td>Seth et al [70]</td>
<td>Devised by 3 fellows of the Royal Australasian College of Surgeons, with experience in performing rhinoplasty and expertise in facial reconstructive surgery</td>
<td>Narrow involving technical aspects of rhinoplasty</td>
<td>Not clear</td>
</tr>
<tr>
<td>Suthar et al [71]</td>
<td>Quizzes from the Case of the Month feature of the American Journal of Neuroradiology</td>
<td>Narrow involving radiology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Walker et al [72]</td>
<td>Devised based on the Global Burden of Disease tool</td>
<td>Narrow involving benign and malignant hepatopancreaticobiliary-related conditions</td>
<td>Not clear</td>
</tr>
<tr>
<td>Wang et al [73]</td>
<td>Medical Exam Help. A total of 10 inpatient and 10 outpatient medical records to form a collection of Chinese medical records after desensitization</td>
<td>Clinical medicine, basic medicine, medical humanities, and relevant laws</td>
<td>Not clear</td>
</tr>
<tr>
<td>Wang et al [74]</td>
<td>The Taiwanese Senior Professional and Technical Examinations for Pharmacists downloaded from the Ministry of Examination website</td>
<td>Broad involving pharmacology and pharmaceutical chemistry, pharmaceutical analysis and pharmacognosy (including Chinese medicine), pharmaceutics and biopharmaceutics, dispensing pharmacy and clinical pharmacy, therapeutics, pharmacy administration, and pharmacy law</td>
<td>Not clear</td>
</tr>
<tr>
<td>Zhou et al [75]</td>
<td>A single clinical case from OrthoBullets, a global clinical collaboration platform for orthopedic surgeons; Written permission to use their clinical case report</td>
<td>Very narrow involving a single orthopedic case</td>
<td>Not clear</td>
</tr>
</tbody>
</table>

**Examples of Optimal Reporting of Each Criterion Within the METRICS Checklist**

The records with the highest scores for each METRICS item, as determined by the average subjective interrater assessments, are shown in Table 4.
Table 4. Included records that had the highest METRICS score per item.

<table>
<thead>
<tr>
<th>Item</th>
<th>Issues considered in each item</th>
<th>Excellent or very good reporting examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Model</td>
<td>What is the model of the generative AI tool used for generating content, and what are the exact settings for each tool?</td>
<td>Baglivo et al [47]: Bing, ChatGPT, Chatsonic, Bard, and YouChat, with full details of the mode and large language model, including plugins</td>
</tr>
<tr>
<td>#2 Evaluation</td>
<td>What is the exact approach used to evaluate the content generated by the generative AI-based model and is it an objective or subjective evaluation?</td>
<td>Al-Ashwal et al [42]: Objective via 2 different clinical reference tools; Alfertshofer et al [43]: Objective based on key answers, with the questions screened independently by 4 investigators; Ali et al [44]: Objective for multiple-choice questions and true or false questions, and subjective for short-answer and essay questions; Aljindan et al [45]: Objective based on key answers and historical performance metrics; and Baglivo et al [47]: Objective based on key answers and comparison with 5th year medical students' performance</td>
</tr>
<tr>
<td>#3a Timing</td>
<td>When is the generative AI model tested exactly and what are the duration and timing of testing?</td>
<td>Baglivo et al [47]; Biswas et al [48]; Fuchs et al [51]; Ghosh &amp; Bir [52]; Hoch et al [58]; Juhi et al [59]; Kumari et al [61]; Kung et al [62]; Oca et al [66]; Pugliese et al [68]; Sallam et al [69]; Wang et al [73]; and Zhou et al [75]</td>
</tr>
<tr>
<td>#3b Transparency</td>
<td>How transparent are the data sources used to generate queries for the generative AI-based model?</td>
<td>Alfertshofer et al [43]</td>
</tr>
<tr>
<td>#4a Range</td>
<td>What is the range of topics tested and are they intersubject or intrasubject with variability in different subjects?</td>
<td>Ali et al [44]; Chen et al [49]; Hoch et al [58]; and Wang et al [73]</td>
</tr>
<tr>
<td>#4b Randomization</td>
<td>Was the process of selecting the topics to be tested on the generative AI-based model randomized?</td>
<td>Alfertshofer et al [43] and Aljindan et al [45]</td>
</tr>
<tr>
<td>#5 Individual</td>
<td>Is there any individual subjective involvement in generative AI content evaluation? If so, did the authors describe the details in full?</td>
<td>Ali et al [44] and Moise et al [65]</td>
</tr>
<tr>
<td>#6 Count</td>
<td>What is the count of queries executed (sample size)?</td>
<td>Alfertshofer et al [43]; Chen et al [49]; Guerra et al [56]; Hoch et al [58]; and Oztermeli &amp; Oztermeli [67]</td>
</tr>
<tr>
<td>#7 Specificity of the prompt or language</td>
<td>How specific are the exact prompts used? Were those exact prompts provided fully? Did the authors consider the feedback and learning loops? How specific are the language and cultural issues considered in the generative AI model?</td>
<td>Alfertshofer et al [43]; Biswas et al [48]; Fuchs et al [51]; Grewal et al [55]; Wang et al [73]; Moise et al [65]; and Pugliese et al [68]</td>
</tr>
</tbody>
</table>

**AI:** artificial intelligence.

**Interrater Assessment of the Included Records Based on METRICS Scores**

The overall mean METRICS score was 3.0 (SD 0.58). For each item, the $\kappa$ interrater reliability ranged from 0.558 to 0.962 ($P<.001$ for the 9 tested items), indicating good to excellent agreement (Table 5).
Table 5. Interrater reliability per METRICS item.

<table>
<thead>
<tr>
<th>METRICS(^a) item</th>
<th>Score Mean(^b) (SD)</th>
<th>Range</th>
<th>Quality</th>
<th>Cohen (\kappa)</th>
<th>Asymptotic standard error</th>
<th>Approximate (T)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3.72 (0.58)</td>
<td>2.5-5.0</td>
<td>Very good</td>
<td>0.820</td>
<td>0.090</td>
<td>6.044</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Timing</td>
<td>2.90 (1.93)</td>
<td>1.0-5.0</td>
<td>Good</td>
<td>0.853</td>
<td>0.076</td>
<td>6.565</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Count</td>
<td>3.04 (1.32)</td>
<td>1.0-5.0</td>
<td>Good</td>
<td>0.962</td>
<td>0.037</td>
<td>10.675</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specificity of prompts and</td>
<td>3.44 (1.25)</td>
<td>1.0-5.0</td>
<td>Very good</td>
<td>0.765</td>
<td>0.086</td>
<td>8.083</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>3.31 (1.16)</td>
<td>1.0-5.0</td>
<td>Good</td>
<td>0.885</td>
<td>0.063</td>
<td>9.668</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Individual factors</td>
<td>2.50 (1.42)</td>
<td>1.0-5.0</td>
<td>Satisfactory</td>
<td>0.865</td>
<td>0.087</td>
<td>6.860</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Transparency</td>
<td>3.24 (1.01)</td>
<td>1.0-5.0</td>
<td>Good</td>
<td>0.558</td>
<td>0.112</td>
<td>5.375</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Range</td>
<td>3.24 (1.07)</td>
<td>2.0-5.0</td>
<td>Good</td>
<td>0.836</td>
<td>0.076</td>
<td>8.102</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Randomization</td>
<td>1.31 (0.87)</td>
<td>1.0-4.0</td>
<td>Suboptimal</td>
<td>0.728</td>
<td>0.135</td>
<td>5.987</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Overall</td>
<td>3.01 (0.58)</td>
<td>1.5-4.1</td>
<td>Good</td>
<td>0.381</td>
<td>0.086</td>
<td>10.093</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\)METRICS: Model, Evaluation, Timing, Range, Randomization, Individual factors, Count, and Specificity of prompts and language.
\(^b\)The mean scores represent the results of evaluating the included studies averaged for the 2 rater scores.

Discussion

Principal Findings

The interpretation and synthesis of credible scientific evidence based on studies evaluating commonly used generative AI-based conversational models (eg, ChatGPT, Bing, and Bard) can be challenging. This is related to the discernible variability in the methods used for the evaluation of such models, as well as the varying styles of reporting. Such variability is fathomable considering the emerging nature of this research field with less than a year of reporting at the time of writing. Therefore, a standardized framework to guide the design of such studies and to delineate the best reporting practices can be beneficial, since rigorous methodology and clear reporting of findings are key attributes of science to reach reliable conclusions with real-world implications.

In this study, a preliminary checklist referred to as “METRICS” was formulated, which can help researchers aspiring to test the performance of generative AI-based models in various aspects of health care education and practice. It is crucial to explicitly state that the proposed METRICS checklist in this study cannot be claimed to be comprehensive or flawless. Nevertheless, this checklist could provide a solid base for future studies and much needed efforts aiming to standardize the reporting of AI-based studies in health care.

The principal finding of this study was the establishment of 9 key themes that are recommended to be considered in the design, testing, and reporting of generative AI-based models in research, particularly in the health care domain. These features included the model of the generative AI tool, evaluation approach, timing of testing and transparency, range of topics tested and randomization of queries, individual factors in the design and assessment, count of queries, and specific prompts and languages used. The relevance of these themes in the design and reporting of generative AI-based model content testing can be illustrated as follows.

First, the variability in generative AI model types used to conduct queries and the variability in settings pose significant challenges for cross-study comparisons. The significant impact of generative AI models on the resultant output is related to the distinct architectures and capabilities of various generative AI models, with expected variability in the performance and quality of the content generated by generative AI [76]. Additionally, various options to configure the models further affect the content generated by generative AI. Consequently, it is important to consider these variations when evaluating research using different generative AI models [77-81]. These issues can be illustrated clearly in the records included in this study that conducted contemporary analysis of at least two models. For example, Al-Ashwal et al [42] showed that Bing had the highest accuracy and specificity for predicting drug-drug interaction, outperforming Bard, ChatGPT-3.5, and ChatGPT-4. Moreover, Baglivo et al [47] showed not only intermodel variability but also intramodel variability in performance in the domain of public health. Additionally, in the context of providing information on rhinoplasty, Seth et al [70] showed that this intermodel variability in performance was the most comprehensible with the content of Bard, followed by ChatGPT and Bing.

Second, the continuous updating of generative AI models introduces significant temporal variability, which would influence the comparability of studies conducted at different times. The updates of generative AI models enhance their capabilities and performance [82]. Consequently, this temporal variability can lead to inconsistencies in synthesizing evidence, as the same model may demonstrate different outputs over time. Therefore, when analyzing or comparing studies involving AI models, it is crucial to consider the specific version and state of the model at the time of each study to accurately interpret and compare results. In this context, it is important to conduct
future longitudinal studies to discern the exact effect of changes in performance of commonly used generative AI models over time.

Third, the count of queries in the evaluation of generative AI models was identified among the pertinent themes of assessment. This appears understandable since studies employing a larger number of queries can provide a more comprehensive evaluation of the tested model. An extensive number of queries can reveal minor weaknesses, despite the difficulty to establish what constitutes an “extensive” number of queries of the minimum number of queries needed to reveal the real performance of a generative AI model in a particular topic. In this study, the number of queries varied from a single case to more than 2500 questions, showing the need for standardization and establishing a clear guide on the number of queries deemed suitable [58,75].

Fourth, a key theme identified in this study was the nature and language of the prompts used to conduct the studies. This involved the imperative of explicitly stating the used prompts and the language in which they were framed. The exact prompting approach and the presence of cultural and linguistic biases appear to be critical factors that can influence the quality of content generated by generative AI-based models [83]. Slight differences in wording or context in the prompt used to generate the generative AI content can lead to recognizable differences in the content generated [36,84,85]. Additionally, feedback mechanisms and learning loops that allow generative AI-based models to learn from interactions can change the model performance for the same query, which might not be consistently accounted for in all studies. These minor variations in prompts across different studies can also complicate the synthesis of evidence, highlighting the need for standardizing such an aspect. Additionally, as highlighted above, generative AI-based models may exhibit biases based on their training data, affecting performance across various cultural or linguistic contexts [86-88]. Thus, studies conducted in different regions or involving various languages might yield varying results. In this study, we found that a majority of the included records tested generative AI-based models using the English language, highlighting the need for more studies on other languages to reveal the possible variability in performance based on language. Comparative studies involving multiple languages can reveal such inconsistencies, for example, the study by Wang et al [74].

In the aforementioned study assessing ChatGPT performance in the Taiwanese pharmacist licensing exam, the performance in the English test was better than that in the Chinese test across all tested subjects [74]. Another comprehensive study by Alfertshofer et al [43] that assessed the performance of ChatGPT on 6 different national medical licensing exams highlighted the variability in performance per country and language. Based on the previous points, more studies that encompass diverse language and cultural perspectives are essential to assess and address possible cultural and language biases in generative AI-based models. Additionally, the design of generative AI-based models trained on a more diverse set of languages and cultural contexts is important to ensure that the training data sets are representative of different linguistic groups, which is of paramount importance in health care. Furthermore, cross-cultural validation studies appear valuable for understanding the performance of generative AI-based models in various language and cultural settings. These approaches could enhance the broad applicability of generative AI-based models in health care to ensure the fair distribution of the benefits of generative AI technologies.

Fifth, an important theme highlighted in this study was the approach of evaluating the content generated by generative AI-based models. Variable methods of assessment can introduce a discernible methodological variability. Specifically, the use of objective assessment ensures consistency in assessment. On the other hand, subjective assessment, even by experts, can vary despite the professional judgment and deep understanding provided by such an expert opinion [89]. Similarly, the number and expertise of evaluators or raters involved in constructing and evaluating the generative AI-based studies were identified as a pertinent theme in this study [90,91]. Variations in rater numbers across studies can lead to inconsistencies in synthesized evidence [68,69,92]. Additionally, the method used to establish agreement (eg, kappa statistics and consensus meetings) might differ in various studies, affecting the comparability of results.

Finally, data-pertinent issues were identified as key themes in this study. These involved the need for full transparency regarding the sources of data used to create the queries (eg, question banks, credible national and international guidelines, clinical reports, etc) [93,94]. Additionally, ethical considerations, such as consent to use copyrighted material and consent or anonymization of the clinical data, should be carefully stated in the evaluation studies of AI-based models. An important aspect that appeared suboptimal in the majority of included records was randomization to reduce or eliminate potential bias in query selection. Thus, this important issue should be addressed in future studies to allow unbiased evaluation of the content of generative AI-based models. Another important aspect is the need to carefully select the topics to be tested, which can belong to a broad domain (eg, medical examination) or a narrow domain (eg, a particular specialty) [95-97]. A comprehensive description of topics is essential to reveal subtle differences in generative AI performance across various domains. Biased query coverage per topic may result in unreliable conclusions regarding generative AI model performance.

The value of the METRICS checklist in guiding researchers toward more rigorous design and transparent reporting of health care studies assessing the performance of generative AI-based models can be highlighted as follows. Two studies exemplified the practical utility of the METRICS checklist (presented in its preprint form) across different research scenarios (health care education and health care practice) [98]. The first study conducted a detailed assessment of ChatGPT-3.5 performance in medical microbiology multiple choice questions compared with dental students [99]. By applying the METRICS checklist retrospectively, the study quality was delineated, including the identification of potential limitations, such as the absence of randomization, thus offering a more critical evaluation of the research design [99].
The second study investigating the performance of both ChatGPT-3.5 and ChatGPT-4 in the context of diagnostic microbiology case scenarios was conceived based on the METRICS checklist [100]. The prospective incorporation of the METRICS checklist was particularly instrumental in refining the study design and the reporting of results [100].

Thus, the aforementioned studies emphasize the effectiveness of the METRICS checklist as a tool to standardize research practice in a rapidly growing research field. The real-world application of the METRICS checklist has been valuable in identifying potential research limitations and in enhancing the overall structure and clarity of the reporting of results. These studies also demonstrate the value of the METRICS checklist for guiding researchers toward more rigorous design and transparent reporting of generative AI-based studies in health care [99,100].

**Limitations**

It is crucial to explicitly mention the need for careful interpretation of the findings based on the following limitations. First, the initial search process involved the broad term “artificial intelligence” and was conducted by a single author, which may have inadvertently resulted in missing relevant references. The record selection process was further limited by the screening of record titles for only 3 generative AI models (ChatGPT, Bing, and Bard) using the EndNote search function. Additionally, the reliance on including published English records, indexed in Scopus, PubMed, or Google Scholar, could raise concerns about potential selection bias and the exclusion of relevant studies. However, it is important to consider this limitation in light of the context of our study, which represents a preliminary report that needs to be validated by future comprehensive and exhaustive studies. Second, it is important to acknowledge that a few pertinent themes could have been overlooked despite our attempt to achieve a thorough analysis, given the limited number of authors. Therefore, future follow-up studies can benefit from inclusion of authors with diverse backgrounds, including different health care disciplines, computer scientists, researchers in the field of human-AI interaction, and AI developers. Additionally, the subjective nature of the pertinent theme selection can be considered as another important caveat in this study. This shortcoming extended to involve the raters’ subjective assessments in assigning different METRICS scores. Moreover, the equal weight given to each item of the checklist in the METRICS score might not be a suitable approach, given the possibility of varying importance of each component. Thus, future comprehensive studies should focus on the relative importance of each METRICS component and its possible impact on the reporting of generative AI model performance. Third, the focus on a few specific generative AI-based conversational models (ie, ChatGPT, Bing, and Bard) can potentially overlook the nuanced aspects of other generative AI models. Nevertheless, our approach was justified by the popularity and widespread use of these particular generative AI-based models. However, it is important for future studies to expand the scope to include models, such as Llama or Claude, which could provide a more comprehensive evaluation of the utility of the METRICS checklist. Lastly, we fully and unequivocally acknowledge that the METRICS checklist is preliminary and needs further verification to ensure its valid applicability.

**Future Perspectives**

The METRICS checklist proposed in this study could be a helpful step toward establishing useful guidelines to design and report the findings of generative AI-based studies. The integration of generative AI models in health care education and practice necessitates a collaborative approach involving health care professionals, researchers, and AI developers. Synthesis of evidence with critical appraisal of the quality of each element in the METRICS checklist is recommended for continuous enhancement of AI output, which would result in successful implementation of AI models in health care while avoiding possible concerns. Regular multidisciplinary efforts and iterative revisions are recommended to ensure that the METRICS checklist properly reflects its original intended purpose of improving the quality of study design and result reporting in this swiftly evolving research field. Future studies would benefit from expanding the scope of literature review and data inclusion, with the incorporation of a wider range of databases, languages, and AI models. This is crucial for reaching the ultimate aim of standardization of the design and reporting of generative AI-based studies.

**Conclusions**

The newly devised “METRICS” checklist may represent a key initial step to motivate the standardization of reporting of generative AI-based studies in health care education and practice. Additionally, the establishment of this algorithm can motivate collaborative efforts to develop universally accepted reporting guidelines for the design and reporting of results of generative AI-based studies. In turn, this can enhance the comparability and reliability of evidence synthesis from these studies. The METRICS checklist, as presented by the findings of this study, can help to elucidate the strengths and limitations of the content generated by generative AI-based models, guiding their future development and application. The standardization offered by the METRICS checklist can be important to ensure the reporting of reliable and replicable results. Subsequently, this can result in the exploitation of the promising potential of generative AI-based models in health care while avoiding its possible concerns. The METRICS checklist could mark the significant progress in the evolving research field. Nevertheless, there is room for refinement through revisions and updates to verify its validity.

**Data Availability**

The data used in this study are available upon request from the corresponding author (Malik Sallam).
Authors' Contributions
Malik Sallam contributed to conceptualization; Malik Sallam, MB, and Mohammed Sallam contributed to methodology; Malik Sallam, MB, and Mohammed Sallam contributed to formal analysis; Malik Sallam, MB, and Mohammed Sallam performed the investigation; Malik Sallam, MB, and Mohammed Sallam contributed to data curation; Malik Sallam contributed to writing—original draft preparation; Malik Sallam, MB, and Mohammed Sallam contributed to writing—review and editing; Malik Sallam contributed to visualization; Malik Sallam contributed to supervision; and Malik Sallam contributed to project administration. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

References


37. Eysenbach G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. JIMIR Med Educ 2023 Mar 06;9:e46885 [FREE Full text] [doi: 10.2196/46885] [Medline: 36863937]


41. Harzing AW. The Publish or Perish Book: Your guide to effective and responsible citation analysis. Melbourne, Australia: Tarma Software Research; 2010.


43. CASP Qualitative Studies Checklist. Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2023-11-10]


Sallam et al

INTERACTIVE JOURNAL OF MEDICAL RESEARCH


Abbreviations

AI: artificial intelligence
CASP: Critical Appraisal Skills Programme
METRICS: Model, Evaluation, Timing, Range/Randoization, Individual factors, Count, and Specificity of prompts and language
A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence–Based Models in Health Care Education and Practice: Development Study Involving a Literature Review

Please cite as:
Sallam M, Barakat M, Sallam M

URL: https://www.i-jmr.org/2024/1/e54704
doi: 10.2196/54704
PMID: 38276872

©Malik Sallam, Muna Barakat, Mohammed Sallam. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 15.02.2024. 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.
Research Trends and Evolution in Radiogenomics (2005-2023): Bibliometric Analysis

Meng Wang¹, MM; Yun Peng², MM; Ya Wang², BMed; Dehong Luo¹, MD

¹Department of Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China
²Department of Radiology, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, China

Corresponding Author:
Dehong Luo, MD
Department of Radiology
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital
Chinese Academy of Medical Sciences and Peking Union Medical College
No.113 Baoheng Road, Longgang Distribution
Shenzhen, 518116
China
Phone: 86 1 391 001 6309
Email: cjr.luodehong@vip.163.com

Abstract

Background: Radiogenomics is an emerging technology that integrates genomics and medical image–based radiomics, which is considered a promising approach toward achieving precision medicine.

Objective: The aim of this study was to quantitatively analyze the research status, dynamic trends, and evolutionary trajectory in the radiogenomics field using bibliometric methods.

Methods: The relevant literature published up to 2023 was retrieved from the Web of Science Core Collection. Excel was used to analyze the annual publication trend. VOSviewer was used for constructing the keywords co-occurrence network and the collaboration networks among countries and institutions. CiteSpace was used for citation keywords burst analysis and visualizing the references timeline.

Results: A total of 3237 papers were included and exported in plain-text format. The annual number of publications showed an increasing annual trend. China and the United States have published the most papers in this field, with the highest number of citations in the United States and the highest average number per item in the Netherlands. Keywords burst analysis revealed that several keywords, including “big data,” “magnetic resonance spectroscopy,” “renal cell carcinoma,” “stage,” and “temozolomide,” experienced a citation burst in recent years. The timeline views demonstrated that the references can be categorized into 8 clusters: lower-grade glioma, lung cancer histology, lung adenocarcinoma, breast cancer, radiation-induced lung injury, epidermal growth factor receptor mutation, late radiotherapy toxicity, and artificial intelligence.

Conclusions: The field of radiogenomics is attracting increasing attention from researchers worldwide, with the United States and the Netherlands being the most influential countries. Exploration of artificial intelligence methods based on big data to predict the response of tumors to various treatment methods represents a hot spot research topic in this field at present.


KEYWORDS
bibliometric; radiogenomics; multiomics; genomics; radiomics

Introduction

Radiogenomics is an emerging technology that combines radiomics and genomics, with the ultimate goal of improving prognosis and outcomes [1]. Radiogenomics can be used to investigate the relationship between imaging features and gene mutations and expression patterns [2-4]. Unlike traditional gene sequencing methods, which are associated with inherent drawbacks such as invasive and high-cost procedures, radiogenomics provides a noninvasive, convenient, and cost-effective method by using quantitative imaging parameters extracted from the entire lesions [5,6]. Many scholars have demonstrated that radiogenomics may predict the pathologic
type, prognosis, and outcome of cancers, including lung cancer and liver cancer, based on pretreatment multimodal imaging (computed tomography [CT] or magnetic resonance imaging [MRI]) [7-10]. This technology has also been proposed as a useful biomarker for nontumor diseases, such as in the diagnosis, classification, and prognostic assessment of coronary heart disease [11].

Radiogenomics is an important potential tool for precision medicine. Some review articles summarized the routine process of radiogenomics and its various applications in the management of disease [12,13]. However, these reviews generally focused on presenting the research directions rather than analyzing the dynamic changes in the field, only highlighting the process and application status of radiogenomics. Bibliometrics can be used to quantitatively analyze the countries, institutions, authors, keywords, and other information related to the entire body of literature in a specific field. This approach can also help to visually display the dynamic progress in the field through network mapping [14]. Therefore, the aim of this study was to summarize the research status and dynamic changes of research hot spots in radiogenomics over time using bibliometric methods, thus providing a comprehensive understanding of the emerging trends in this field.

Methods

Bibliometric Data Acquisition

The published literature on radiogenomics was retrieved from the Web of Science Core Collection (WoSCC), which is the most widely used database in bibliometric analysis, on March 1, 2024 [15]. The initial search phase showed that the first relevant article in this field was published in 2005; hence, we restricted the publication time period to 2005-2023 [16]. The search string was as follows: "(TS=(Radiogenomics) OR TS=(Radiomics AND genomics) OR TS=((Radiomics) AND (gene* OR DNA OR RNA OR expression OR mutation OR molecular subtype))) AND FPY=(2005-2023)."

The literature retrieval and refining processes were carried out by one author (YW), while the other authors supervised the whole process. A total of 3669 documents were retrieved. The refine function of the WoSCC website was used to screen the relevant article in this field, only highlighting the process and application status of radiogenomics. Certain data, including the number of publications, impact factor, h-index, and average per item were used in the country, institution, and authors, were retrieved from the WoSCC website. The h-index and average per item were used in the country, institution, and authors.

Synonym Substitution of Keywords

The keywords with the same meaning were merged by synonym substitution. For example, terms such as “computed tomography,” “computed tomography (ct),” “computed-tomography,” “ct,” and “ct images” were uniformly labeled as “CT.” The full list of keyword synonyms is provided in Multimedia Appendix 2. Figure 1 shows a workflow of the analytical procedures.
Results

Annual Publications Trend

A total of 3237 papers were included for the final analysis. No duplicate article was found. The results indicated a clear upward trend in research on radiogenomics since 2013 (Figure 2).

Journals

Table 1 presents the top 15 journals with the highest number of publications on radiogenomics. *Frontiers in Oncology*, *Cancers*, and *European Radiology* were the top three journals publishing in this field with 284, 196, and 135 papers, respectively.
Table 1. Top 15 journals publishing in the field of radiogenomics.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Journal</th>
<th>Articles, n</th>
<th>IF&lt;sup&gt;a&lt;/sup&gt; (2022)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Frontiers in Oncology</td>
<td>284</td>
<td>4.7003</td>
</tr>
<tr>
<td>2</td>
<td>Cancers</td>
<td>196</td>
<td>5.5999</td>
</tr>
<tr>
<td>3</td>
<td>European Radiology</td>
<td>135</td>
<td>5.9003</td>
</tr>
<tr>
<td>4</td>
<td>Scientific Reports</td>
<td>103</td>
<td>4.6</td>
</tr>
<tr>
<td>5</td>
<td>Diagnostics</td>
<td>73</td>
<td>3.5999</td>
</tr>
<tr>
<td>6</td>
<td>Medical Physics</td>
<td>63</td>
<td>3.8001</td>
</tr>
<tr>
<td>7</td>
<td>Journal of Magnetic Resonance Imaging</td>
<td>60</td>
<td>4.3997</td>
</tr>
<tr>
<td>8</td>
<td>European Journal of Radiology</td>
<td>57</td>
<td>3.3</td>
</tr>
<tr>
<td>9</td>
<td>Academic Radiology</td>
<td>50</td>
<td>4.8003</td>
</tr>
<tr>
<td>10</td>
<td>Physics in Medicine And Biology</td>
<td>49</td>
<td>3.5</td>
</tr>
<tr>
<td>11</td>
<td>British Journal of Radiology</td>
<td>45</td>
<td>2.6002</td>
</tr>
<tr>
<td>12</td>
<td>Radiology</td>
<td>41</td>
<td>19.7005</td>
</tr>
<tr>
<td>13</td>
<td>European Journal of Nuclear Medicine and Molecular Imaging</td>
<td>40</td>
<td>9.1005</td>
</tr>
<tr>
<td>14</td>
<td>Abdominal Radiology</td>
<td>34</td>
<td>2.4002</td>
</tr>
<tr>
<td>15</td>
<td>BMC Medical Imaging</td>
<td>33</td>
<td>2.7001</td>
</tr>
</tbody>
</table>

<sup>a</sup>IF: impact factor.

Countries and Institutions

A total of 71 countries have published articles related to radiogenomics. Table 2 highlights the top 10 countries in terms of the number of publications. China ranks first with 1470 articles, followed by the United States with 891 articles and Italy with 326 articles. The United States obtained the highest citation count and the second highest average citation per item. Only three countries, the Netherlands (99.86), the United States (52.11), and Canada (51.46), have average citations per item above 50, with the average citation count for the Netherlands standing high above those of the other countries. Six countries, including the United States, England, Italy, Canada, China, and the Netherlands, had high BC values (≥0.1).

Table 2. Top 10 productive countries with the most publications in the field of radiogenomics.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Total link strength</th>
<th>BC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Country</th>
<th>Publications, n</th>
<th>H-index</th>
<th>Times cited</th>
<th>Average per item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>329</td>
<td>0.11</td>
<td>China</td>
<td>1470</td>
<td>68</td>
<td>25,319</td>
<td>17.22</td>
</tr>
<tr>
<td>2</td>
<td>820</td>
<td>0.36</td>
<td>United States</td>
<td>891</td>
<td>87</td>
<td>46,426</td>
<td>52.11</td>
</tr>
<tr>
<td>3</td>
<td>311</td>
<td>0.13</td>
<td>Italy</td>
<td>326</td>
<td>41</td>
<td>7327</td>
<td>22.48</td>
</tr>
<tr>
<td>4</td>
<td>353</td>
<td>0.06</td>
<td>Germany</td>
<td>233</td>
<td>39</td>
<td>7966</td>
<td>34.19</td>
</tr>
<tr>
<td>5</td>
<td>418</td>
<td>0.19</td>
<td>England</td>
<td>209</td>
<td>37</td>
<td>6185</td>
<td>29.59</td>
</tr>
<tr>
<td>6</td>
<td>328</td>
<td>0.1</td>
<td>Netherlands</td>
<td>161</td>
<td>37</td>
<td>16,077</td>
<td>99.86</td>
</tr>
<tr>
<td>7</td>
<td>54</td>
<td>0.01</td>
<td>South Korea</td>
<td>142</td>
<td>31</td>
<td>4296</td>
<td>30.25</td>
</tr>
<tr>
<td>8</td>
<td>248</td>
<td>0.12</td>
<td>Canada</td>
<td>140</td>
<td>36</td>
<td>7204</td>
<td>51.46</td>
</tr>
<tr>
<td>9</td>
<td>223</td>
<td>0.08</td>
<td>France</td>
<td>128</td>
<td>32</td>
<td>4099</td>
<td>32.02</td>
</tr>
<tr>
<td>10</td>
<td>181</td>
<td>0.01</td>
<td>Switzerland</td>
<td>88</td>
<td>23</td>
<td>1893</td>
<td>21.51</td>
</tr>
</tbody>
</table>

<sup>a</sup>BC: betweenness centrality.

Among the total 71 countries/regions that have contributed to radiogenomics research, 40 have published 5 or more documents. Figure 3A presents the visualization of the countries network. A total of 3523 institutions have contributed to radiogenomics research and 432 institutions published 5 or more documents. Figure 3B presents the visualization of the institutions network. These results show that there is relatively more cooperation between developed countries and their institutions.
Figure 3. Countries (A) and institutions (B) collaboration networks in the field of radiogenomics. The size of the nodes corresponds to the number of published documents and the line width between nodes indicates the strength of coauthorship. Thicker lines indicate a higher frequency of cooperation.

Authors
A total of 17,727 authors have contributed to radiogenomics. Table 3 presents the top 10 productive authors and the most cited authors in this field. The authors with the most publications are Philippe Lambin (33 papers), Catharine M West (32 papers), and Robyn Gillies (32 papers). Three of the top 10 productive authors, including Philippe Lambin (the Netherlands), Robyn Gillies (Australia), and Hugo Aerts (United States), also ranked in the top 3 for citations.

Table 3. Top 10 most productive and highly cited authors on radiogenomics.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Publications</th>
<th>Articles, n</th>
<th>Country</th>
<th>H-index</th>
<th>Average per item</th>
<th>Citations</th>
<th>Citations, n</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Philippe Lambin</td>
<td>33</td>
<td>Netherlands</td>
<td>21</td>
<td>358.06</td>
<td>Philippe Lambin</td>
<td>1573</td>
<td>Netherlands</td>
</tr>
<tr>
<td>2</td>
<td>Catharine M West</td>
<td>32</td>
<td>England</td>
<td>17</td>
<td>44.78</td>
<td>Robyn Gillies</td>
<td>1109</td>
<td>Australia</td>
</tr>
<tr>
<td>3</td>
<td>Robyn Gillies</td>
<td>32</td>
<td>Australia</td>
<td>24</td>
<td>486.13</td>
<td>Hugo Aerts</td>
<td>965</td>
<td>United States</td>
</tr>
<tr>
<td>4</td>
<td>Evis Sala</td>
<td>26</td>
<td>Italy</td>
<td>17</td>
<td>35.92</td>
<td>JIM Van Griethuyser</td>
<td>720</td>
<td>United States</td>
</tr>
<tr>
<td>5</td>
<td>Andre Dekker</td>
<td>25</td>
<td>Netherlands</td>
<td>12</td>
<td>448.4</td>
<td>Alex Zwanenburg</td>
<td>597</td>
<td>Germany</td>
</tr>
<tr>
<td>6</td>
<td>Jie Tian</td>
<td>25</td>
<td>China</td>
<td>14</td>
<td>48.04</td>
<td>Chintan Parmar</td>
<td>448</td>
<td>United States</td>
</tr>
<tr>
<td>7</td>
<td>Sarah L Kerns</td>
<td>25</td>
<td>United States</td>
<td>17</td>
<td>44.72</td>
<td>David N Louis</td>
<td>429</td>
<td>United States</td>
</tr>
<tr>
<td>8</td>
<td>Hugo Aerts</td>
<td>24</td>
<td>United States</td>
<td>21</td>
<td>526.42</td>
<td>Philip Kickingeder</td>
<td>390</td>
<td>Germany</td>
</tr>
<tr>
<td>9</td>
<td>Shao Feng Duan,</td>
<td>24</td>
<td>China</td>
<td>11</td>
<td>12.67</td>
<td>Yan-Qi Huang</td>
<td>382</td>
<td>China</td>
</tr>
<tr>
<td>10</td>
<td>Seung-Koo Lee</td>
<td>22</td>
<td>South Korea</td>
<td>10</td>
<td>23.91</td>
<td>Kumar Vinod</td>
<td>376</td>
<td>India</td>
</tr>
</tbody>
</table>

Keywords
There were 7624 keywords identified in this study and 466 keywords appeared more than 9 times. Figure 4 presents an overlay visualization map of the co-occurring keywords. Table 4 presents the top 30 keywords based on their occurrence frequency. Apart from “radiomics” and “radiogenomics,” the most frequent keyword was “machine learning” (n=779), followed by “CT” (n=580) and “carcinoma” (n=569).
Figure 4. Overlay visualization map of keywords. The node size indicates the frequency of keyword occurrences and the color represents the average publication year of the identified keywords.
Table 4. The top 30 keywords with the highest frequency in the field of radiogenomics.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Keywords</th>
<th>Total link strength</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>radiomics</td>
<td>13,538</td>
<td>1918</td>
</tr>
<tr>
<td>2</td>
<td>machine learning</td>
<td>5816</td>
<td>779</td>
</tr>
<tr>
<td>3</td>
<td>CT^a</td>
<td>4300</td>
<td>580</td>
</tr>
<tr>
<td>4</td>
<td>carcinoma</td>
<td>4250</td>
<td>569</td>
</tr>
<tr>
<td>5</td>
<td>features</td>
<td>4279</td>
<td>538</td>
</tr>
<tr>
<td>6</td>
<td>radiogenomics</td>
<td>4102</td>
<td>535</td>
</tr>
<tr>
<td>7</td>
<td>survival analysis</td>
<td>3984</td>
<td>500</td>
</tr>
<tr>
<td>8</td>
<td>imaging</td>
<td>3650</td>
<td>478</td>
</tr>
<tr>
<td>9</td>
<td>classification</td>
<td>3367</td>
<td>439</td>
</tr>
<tr>
<td>10</td>
<td>MRI^b</td>
<td>3323</td>
<td>423</td>
</tr>
<tr>
<td>11</td>
<td>predictors</td>
<td>3427</td>
<td>416</td>
</tr>
<tr>
<td>12</td>
<td>diagnosis</td>
<td>2247</td>
<td>310</td>
</tr>
<tr>
<td>13</td>
<td>texture analysis</td>
<td>2489</td>
<td>303</td>
</tr>
<tr>
<td>14</td>
<td>deep learning</td>
<td>2076</td>
<td>273</td>
</tr>
<tr>
<td>15</td>
<td>tumors</td>
<td>2095</td>
<td>271</td>
</tr>
<tr>
<td>16</td>
<td>breast cancer</td>
<td>1958</td>
<td>260</td>
</tr>
<tr>
<td>17</td>
<td>artificial intelligence</td>
<td>2074</td>
<td>254</td>
</tr>
<tr>
<td>18</td>
<td>expression</td>
<td>1906</td>
<td>242</td>
</tr>
<tr>
<td>19</td>
<td>biomarkers</td>
<td>1906</td>
<td>240</td>
</tr>
<tr>
<td>20</td>
<td>heterogeneity</td>
<td>1902</td>
<td>227</td>
</tr>
<tr>
<td>21</td>
<td>F-18-FDG PET^c</td>
<td>1874</td>
<td>225</td>
</tr>
<tr>
<td>22</td>
<td>prognosis</td>
<td>1670</td>
<td>219</td>
</tr>
<tr>
<td>23</td>
<td>glioblastoma</td>
<td>1806</td>
<td>212</td>
</tr>
<tr>
<td>24</td>
<td>gliomas</td>
<td>1611</td>
<td>198</td>
</tr>
<tr>
<td>25</td>
<td>lung cancer</td>
<td>1485</td>
<td>193</td>
</tr>
<tr>
<td>26</td>
<td>radiotherapy</td>
<td>1258</td>
<td>175</td>
</tr>
<tr>
<td>27</td>
<td>nomogram</td>
<td>1228</td>
<td>167</td>
</tr>
<tr>
<td>28</td>
<td>models</td>
<td>1083</td>
<td>152</td>
</tr>
<tr>
<td>29</td>
<td>recurrence</td>
<td>1329</td>
<td>152</td>
</tr>
<tr>
<td>30</td>
<td>system</td>
<td>983</td>
<td>144</td>
</tr>
</tbody>
</table>

^aCT: computed tomography.

^bMRI: magnetic resonance imaging.

^cF-18 FDG PET: fludeoxyglucose F18 positron emission tomography.

Keywords Burst and References Cluster

The top 25 keywords with the strongest citation bursts are depicted in Figure 5A. The top 3 keywords with the strongest citation bursts were “gene expression” (17.44), “single nucleotide polymorphism” (16.05), and “genome-wide association” (14.16). The keywords “big data,” “magnetic resonance spectroscopy,” “renal cell carcinoma,” “stage,” and “temozolomide” experienced a citation burst in recent years. Figure 5B illustrates the reference clusters along horizontal timelines. CiteSpace generated 8 clusters: cluster 0 for lower-grade glioma, cluster 1 for lung cancer histology, cluster 2 for lung adenocarcinoma, cluster 3 for breast cancer, cluster 4 for radiation-induced lung injury, cluster 5 for epidermal growth factor receptor (EGFR) mutation, cluster 6 for late radiotherapy toxicity, and cluster 7 for artificial intelligence.
Based on this comprehensive analysis, the time frame from 2005 to 2023 can be artificially segmented into distinct phases based on the evolution of hot topics in the field. The first phase is approximately from 2005 to 2010, represented by the keywords “radiation-induced lung injury,” “late radiotherapy toxicity,” and “single nucleotide polymorphism.” The second phase spans from approximately 2011 to 2017, represented by the keywords “lung cancer histology,” “breast cancer,” “tumor heterogeneity,” “contrast enhanced MRI,” and “F-18 FDG PET” (fluodeoxyglucose F18 positron emission tomography). The third phase is after 2018, represented by the keywords “phenotypes,” “big data,” “magnetic resonance spectroscopy,” “renal cell carcinoma,” “stage,” “EGFR mutation,” “temozolomide,” and “artificial intelligence.”

Dynamic Publication Trend and Evolutionary Trajectory

The time frame of publications in this relatively new field can be artificially segmented into three phases according to the evolution of hot topics. In the first phase (2005 to 2010), radiogenomics primarily focused on the genetic variation associated with the response to radiation therapy in the field of radiation oncology [12]. Radiation therapy plays a crucial role in tumor treatment, accounting for 50% of all tumor therapies performed worldwide [21]. However, individuals with similar tumors often exhibit significant differences in radiosensitivity, and many patients experience various types of adverse reactions, including radiation-induced lung injury and late radiotherapy toxicity, after radiation therapy [22,23]. To develop precise and personalized treatments that achieve the best efficacy with minimal adverse reactions, researchers have been searching for biomarkers that can predict treatment outcomes. Through analysis of the complete genome using techniques such as genome-wide association analysis, particularly focusing on single-nucleotide polymorphism markers, researchers have identified numerous genomic variation sites associated with the response to radiotherapy [24,25].

In the second phase (2011 to 2017), the concept of radiogenomics expanded. Studies incorporating medical imaging features and biological parameters beyond genomics were also included in radiogenomics studies [26]. It is believed that the features from medical images such as MRI, CT, and PET-CT are closely related to tumor heterogeneity. Therefore, researchers have extracted the features (including semantic features and texture analysis features) of the tumors and adopted radiomics for a differentiation diagnosis, such as histological subtype identification. Doshi et al [27] found that MRI-based first-order texture metrics can help discriminate between type 1 and type 2 papillary renal cell carcinoma.

From the late second phase onward (ie, 2017 to 2023), the purpose of radiogenomics is not only limited to the prediction of radiotherapy side effects or differential diagnosis but also to analyzing the relationship between gene expression and imaging data. For example, through the analysis of quantitative features of enhanced MRI, Yeh et al [28] found that partial features were correlated with the expression levels of molecules in the Janus kinase-signal transducer and activator of transcription and...
vascular endothelial growth factor signaling pathways in breast cancer [28].

The distinction between the second and third phases is unclear, with some of the hot topics beginning during the second phase and continuing beyond 2018. In the third phase, the scope of radiogenomics has gradually expanded and become more comprehensive. From the view of raw data, apart from conventional images, some functional imaging techniques such as magnetic resonance spectroscopy have started to be used for radiogenomics analysis [29]. Moreover, with the use of a picture archiving and communication system, the storage and re-extraction of medical data are more convenient, which promotes the progress of big data research and improves the credibility of radiogenomics. From the research purposes perspective, more and more therapeutic methods (eg, neoadjuvant therapy, chemoarterial chemoembolization, transcatheter arterial chemoembolization) have been developed and applied in clinical practice. Researchers are beginning to explore the use of radiogenomics to identify patients who may not be sensitive to certain therapies, thereby reducing unnecessary treatment to avoid side effects [30]. From the view of research methods, the studies in the second phase tended to screen for the quantitative features (which were manually extracted in most cases) associated with gene expression status. At present, many studies use machine learning algorithms that are sometimes combined with deep learning algorithms, which can automatically segment lesions to achieve higher predictive performance [27].

Limitations

Our study has several limitations. First, only research articles and review articles published in English from the WoSCC were included in this analysis, potentially introducing language, publication type, and database biases. Second, this study focused on an in-depth analysis of the dynamic trend and evolutionary trajectory in radiogenomics based on the keywords and references. There are other analyses that could have been considered to better understand the evolution of radiogenomics as a subject, such as more comparative analyses of various factors (ie, authors, countries, keywords, and journals). Third, our results showed that radiogenomics is currently applied mostly in cancer research. Bibliometrics may overlook other topics that are not current research hot spots in the field. For example, keywords related to nononcologic diseases such as mental illness are not included in the tables and figures.

Conclusion

In conclusion, radiogenomics has attracted substantial attention in recent years. The United States and the Netherlands are the leading countries publishing research in this field, obtaining the highest total citations and average per item, respectively. Before 2010, radiogenomics was mainly used to explore the genetic factors associated with radiotherapy-induced toxicity. Subsequently, the field has evolved to encompass the combination of radiomics and genomics, enabling the prediction of cancer histology, gene mutations, and gene expression status based on the tumor heterogeneity information obtained from medical imaging. More and more researchers tend to be exploring the feasibility of radiogenomics to predict the response of tumors to various treatments such as neoadjuvant chemotherapy. The application of artificial intelligence methods based on big data is emerging as a hot spot research topic in this field at present.

Acknowledgments

This work was supported by the Shenzhen High-level Hospital Construction Fund and Shenzhen Clinical Research Center for Cancer (grant (2021)287).

Authors' Contributions

MW performed the formal analysis, validation, investigation, visualization, and data curation. YP and YW contributed to methodology, software, formal analysis, validation, visualization, study supervision, and resources. DL was responsible for study conceptualization, validation, formal analysis, supervision, project administration, and data curation. All authors contributed to the writing and review of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Primary data downloaded as plain text-files from the Web of Science Core Collection (WoSCC) database for bibliometric analysis.

[ZIP File (Zip Archive), 9148 KB - ijmr_v13i1e51347_app1.zip ]

Multimedia Appendix 2
Synonym substitution of keywords.

[DOC File, 76 KB - ijmr_v13i1e51347_app2.doc ]

References

https://www.i-jmr.org/2024/1/e51347


**Abbreviations**

- BC: betweenness centrality
- CT: computed tomography
- EGFR: epidermal growth factor receptor
- F-18 FDG PET: fludeoxyglucose F18 positron emission tomography
- MRI: magnetic resonance imaging
- WoSCC: Web of Science Core Collection
Corrigenda and Addenda

Correction: Influence of Environmental Factors and Genome Diversity on Cumulative COVID-19 Cases in the Highland Region of China: Comparative Correlational Study

Zhuoga Deji1,2*, MSc; Yuantao Tong1*, PhD; Honglian Huang1,3, PhD; Zeyu Zhang1, PhD; Meng Fang4, PhD; M James C Crabbe5,6,7, PhD; Xiaoyan Zhang1, PhD; Ying Wang3, PhD

1Research Center for Translational Medicine, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China
2Information School, The University of Sheffield, Sheffield, United Kingdom
3Department of Clinical Laboratory Medicine Center, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China
4Department of Laboratory Medicine, Shanghai Eastern Hepatobiliary Surgery Hospital, Shanghai, China
5Wolfson College, Oxford University, Oxford, United Kingdom
6Institute of Biomedical and Environmental Science & Technology, University of Bedfordshire, Bedfordshire, United Kingdom
7School of Life Sciences, Shanxi University, Shanxi, China
*these authors contributed equally

Corresponding Author:
Ying Wang, PhD
Department of Clinical Laboratory Medicine Center
Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine
110 Ganhe Rd
Shanghai, 200437
China
Phone: 86 021 65161782
Fax: 86 021 65161782
Email: nadger_wang@139.com

Related Article:
Correction of: https://www.i-jmr.org/2024/1/e43585
doi:10.2196/59070

In "Influence of Environmental Factors and Genome Diversity on Cumulative COVID-19 Cases in the Highland Region of China: Comparative Correlational Study" (Interact J Med Res 2024;13:e43585) the authors noted one error.

In the originally published manuscript, author Zhuoga Deji was noted as having contributed to the manuscript as the first author.

This has been corrected to note that both authors Zhuoga Deji and Yuantao Tong contributed equally to the manuscript, and they should be regarded as joint first authors.

The correction will appear in the online version of the paper on the JMIR Publications Interactive Journal of Medical Research publication website on April 10, 2024, 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.
Cardiovascular Comorbidities in COVID-19: Comprehensive Analysis of Key Topics

Rene Markovič, PhD; Luka Ternar, MD; Tim Trstenjak, MD; Marko Marhl, PhD; Vladimir Grubelnik, PhD

1Faculty of Natural Sciences and Mathematics, University of Maribor, Maribor, Slovenia
2Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia
3Faculty of Medicine, University of Maribor, Maribor, Slovenia
4Faculty of Education, University of Maribor, Maribor, Slovenia

Corresponding Author:
Vladimir Grubelnik, PhD
Faculty of Electrical Engineering and Computer Science
University of Maribor
Koroška cesta 46
Maribor, SI-2000
Slovenia
Phone: 386 2 220 73 96
Email: vlado.grubelnik@um.si

Abstract

Background: The interrelation between COVID-19 and various cardiovascular and metabolic disorders has been a critical area of study. There is a growing need to understand how comorbidities such as cardiovascular diseases (CVDs) and metabolic disorders affect the risk and severity of COVID-19.

Objective: The objective of this study is to systematically analyze the association between COVID-19 and cardiovascular and metabolic disorders. The focus is on comorbidity, examining the roles of CVDs such as embolism, thrombosis, hypertension, and heart failure, as well as metabolic disorders such as disorders of glucose and iron metabolism.

Methods: Our study involved a systematic search in PubMed for literature published from 2000 to 2022. We established 2 databases: one for COVID-19–related articles and another for CVD-related articles, ensuring all were peer-reviewed. In terms of data analysis, statistical methods were applied to compare the frequency and relevance of MeSH (Medical Subject Headings) terms between the 2 databases. This involved analyzing the differences and ratios in the usage of these terms and employing statistical tests to determine their significance in relation to key CVDs within the COVID-19 research context.

Results: The study revealed that “Cardiovascular Diseases” and “Nutritional and Metabolic Diseases” were highly relevant as level 1 Medical Subject Headings descriptors in COVID-19 comorbidity research. Detailed analysis at level 2 and level 3 showed “Vascular Disease” and “Heart Disease” as prominent descriptors under CVDs. Significantly, “Glucose Metabolism Disorders” were frequently associated with COVID-19 comorbidities such as embolism, thrombosis, and heart failure. Furthermore, iron deficiency (ID) was notably different in its occurrence between COVID-19 and CVD articles, underlining its significance in the context of COVID-19 comorbidities. Statistical analysis underscored these differences, highlighting the importance of both glucose and iron metabolism disorders in COVID-19 research.

Conclusions: This work lays the foundation for future research that utilizes a knowledge-based approach to elucidate the intricate relationships between these conditions, aiming to develop more effective health care strategies and interventions in the face of ongoing pandemic challenges.

(KEYWORDS COVID-19; cardiovascular diseases; metabolic disorders; embolism and thrombosis; hypertension; hyperglycemia; iron metabolism disorders; MeSH; embolism; thrombosis; heart failure; nutritional; vascular disease; glucose; effective)

**Introduction**

The SARS-CoV-2 virus, which causes the disease COVID-19, has impacted all areas of our lives. The scientific community has shown an unprecedented and coordinated response to this global pandemic [1,2]. This has led to a rapid acquisition of new knowledge in a wide range of scientific fields and, simultaneously, to new questions that needed to be answered [3-5]. Some of the most important questions concern the origins and causes leading to the severe form of COVID-19 or even death [6,7]. One of the most important pathological aspects of COVID-19 disease is its impact on the cardiovascular system, more specifically cardiovascular disease (CVD) [8-11]. The link between COVID-19 and CVD has been demonstrated and confirmed in numerous studies. A recent scientific review article by Vosko et al [12] offers an extensive overview of the literature on the interaction between COVID-19 and CVDs. The authors describe how COVID-19 can act as a yet unrecognized risk modifier for CVD, including risk factors such as diabetes mellitus [13] or arterial hypertension [14]. In the study by Vosko et al [12], an increased incidence of CVD and poorer clinical outcomes were observed in individuals with preexisting CVD, noting conditions like myocarditis, acute coronary syndrome, heart failure, thromboembolic complications, and arrhythmias. Furthermore, the article by Vosko et al [12] summarizes the mechanisms through which COVID-19 can affect CVD, including the impact on endothelial cells and inflammation, which can increase the risk for atherosclerosis and other cardiovascular events. Additionally, a review study [15] was conducted to demonstrate the connection between COVID-19 and CVD. This study provides a detailed examination of the impact of COVID-19 on different cells in myocardial tissue and offers an overview of the clinical manifestations of cardiovascular involvement in the pandemic.

The most striking link between COVID-19 and CVD involves the angiotensin-converting enzyme 2 (ACE2), which is the main receptor for the glycoprotein membrane spike of SARS-CoV-2 [16-18]. ACE2 is bound to cell membranes in various tissues of the vascular system [19]. Considering its importance in CVD, a population-based study showed that higher ACE2 plasma levels are associated with a greater risk of severe CVD [20]. COVID-19 has been found to increase the risk of cardiogenic shock [21,22], cardiac arrhythmias [23,24], acute myocardial injury [25,26], and sometimes sudden death in patients with CVD [15,27,28], and at the same time, patients with CVD have a higher risk of mortality due to COVID-19.

ACE2 is an important down-regulator of the renin-angiotensin-aldosterone system (RAAS), which plays a significant role in controlling arterial blood pressure [29]. Various studies have investigated the dysregulation of ACE2 in different cells of patients with CVD, indicating an involvement of the RAAS [30,31]. For example, downregulation was found primarily in fibroblasts and the vascular smooth muscle of veins with dilated or hypertrophic cardiomyopathy [32,33]. Conversely, an upregulation of ACE2 is mainly observed in the cardiomyocytes of patients with ischemic and non-ischemic cardiomyopathy [32-34]. It is also noted in the lungs of patients with hypertension, cerebrovascular disease, coronary artery disease, and other comorbidities such as diabetes [35], which may be attributable to the joint treatment of such comorbidities in addition to the disease itself [36]. This correlation is further supported by biochemical and genetic analyses, as patients with heart failure show increased ACE2 expression. ACE is found in 7.6% of all heart cells, compared to only 5.88% in healthy individuals. This is even more pronounced in cardiomyocytes, where 9.87% of all cardiomyocytes in heart failure express ACE2, whereas in healthy hearts, the figure is 6.75% of cardiomyocytes. This is reversed in arterial vascular cells: heart failure shows positive ACE2 expression in 7.93% of vascular cells and 19.4% in healthy individuals [37]. The invasion of SARS-CoV-2 upregulates the activity of the protease ADAM17, which in turn downregulates ACE2 by cleaving it from the cell surface. This process, known as “shedding,” and is very important for understanding the cardiovascular effects of COVID-19. Recognizing the beneficial effects of Ang-1-7 signaling, we understand that disruption of this pathway through shedding leads to the predominance of the RAAS, causing hypertension, fibrotic remodeling, inflammation, and sodium retention [38,39].

Novel big data streams have created interesting opportunities to synthesize research and identify hotspots of big data in infectious disease epidemiology [40]. Furthermore, big data bibliometric analyses can reveal trends and project future developments in each scientific discipline [41-43]. Thus, based on bibliometric analysis, a study was conducted [44], that aimed to investigate the international scientific output on the relationship between COVID-19 and CVDs. The findings revealed that the United States and China are at the forefront in both the quantity and quality of publications in this area. Additionally, the analysis indicated that researchers have paid special attention to cardiovascular comorbidities, outcomes, and regenerative medicine in the context of COVID-19. Such innovative analytical approaches, which leverage extensive big data resources, are particularly crucial for deciphering the complex dynamics of comorbidity patterns observed in COVID-19 and CVDs. By integrating big data insights with traditional epidemiological methods, our study not only contributes to a deeper understanding of these comorbidities but also opens new avenues for predictive analytics in health care.

Considering all this evidence, a critical interface between the virus and CVD has emerged, posing unique challenges to health care systems worldwide. This study aims to unravel the complex relationship between COVID-19 and CVD, addressing a significant gap in our current understanding of the comorbidity dynamics of these diseases. Utilizing a novel approach with MeSH (Medical Subject Headings) descriptors, we systematically analyze a wide range of literature to identify key patterns and themes. Our study not only sheds light on the increased risks and outcomes associated with these comorbidities, but also paves the way for future research methods. This manuscript is organized to first explain our methodological approach, followed by a presentation of our findings, a discussion of their implications, and concludes with insights that have the potential to inform future health care strategies and interventions.
Methods

Overview
We conducted a search on PubMed [45] with specific search queries on CVD and SARS-CoV-2 and limited our search to articles published from the year 2000 onwards. Between January 1, 2000, and September 30, 2021, we collected all relevant entries in PubMed. From these entries, we selected only peer-reviewed scientific publications. We then created 2 databases: one for articles related to COVID-19 (the COVID-19 database) and another for articles related to CVDs (the CVD database). The databases had a similar organization and stored 2 primary pieces of information: the PubMed identifier (PMID) and all the MeSH descriptors provided for an article. The following sections describe in detail the creation of the databases, the use of the MeSH classification scheme, and the analyses performed.

MeSH Classification
The MeSH thesaurus is a controlled and hierarchically organized vocabulary developed and curated by the National Library of Medicine (NLM). The assignment of MeSH descriptors to papers by professional indexers at the NLM is highly consistent and an efficient method for describing the main topics of an article. Consequently, the MeSH classification system offers an organized approach for sorting and accessing medical knowledge. This knowledge is represented by MeSH descriptors or MeSH terms, which are organized hierarchically to facilitate efficient retrieval of biomedical and health-related information from the NLM databases.

In the MeSH tree, the 16 main categories form the foundation of its hierarchical structure. Each main category branches into level 1 (LV1) subbranches, representing more specific aspects of the primary category. These LV1 subbranches further divide into level 2 (LV2) subbranches, offering an even more detailed classification. This pattern continues, with each subsequent level—level 3 (LV3), level 4, and so forth—delving deeper into specialized topics, ensuring a comprehensive and nuanced organization of medical subjects. Overall, the MeSH descriptors are structured hierarchically across 13 levels of subbranches. The coding of MeSH descriptors involves assigning unique alphanumeric identifiers to each descriptor in the MeSH database. These codes serve as precise references, facilitating information retrieval and classification in medical and health-related databases. Typically, MeSH codes consist of a combination of letters and numbers. The letters often represent the main category or aspect of health or medicine the descriptor pertains to, while the numbers provide a unique identifier within that category. Figure 1 presents a schematic representation of the MeSH tree.

For our analysis, we developed a Python script capable of mining relevant publications from PubMed through their API. It extracts the MeSH descriptors associated with an article, along with its unique PMID, and translates a given MeSH descriptor code to the corresponding MeSH descriptor name. Since the code of the MeSH descriptor embeds the location of the term in the MeSH tree, our script can determine the branches from which a MeSH descriptor originates. Our analysis primarily focuses on the “Diseases” main branch (denoted by the letter C), especially the “Cardiovascular Diseases” subcategory (LV1 subbranch C14; Figure 1). The hierarchical structure of the MeSH tree enables an in-depth analysis of topics at different levels of specificity, as illustrated in Figure 1. A more detailed description of the algorithm is given in the subsequent sections.

Creation of the Database
Using Python and the PubMed API, Entrez, our algorithm retrieved relevant information from the PubMed database on COVID-19 and CVDs. We utilized MeSH descriptors as search parameters. For the CVD database, our search query was “Cardiovascular Diseases [MeSH Terms],” while for the COVID-19 database, it was “COVID-19 [MeSH Terms].” Our inclusion criteria were limited to articles from peer-reviewed journals. Figure 2 provides a comprehensive breakdown of the records obtained for each query, categorized by publication type, focusing on the PMID and associated MeSH terms.
For the selected publications, we retrieved raw XML data from PubMed and extracted 2 pieces of information from each XML file: all MeSH descriptors and the PMID. The latter was used to remove duplicate entries. Following this procedure, we created 2 databases: one for the COVID-19 query and another for the CVD query. In addition to these databases, we also developed a graph of MeSH descriptors, which was used for information retrieval.

Data Analysis


Our initial step in the analysis involved calculating the relative frequency of each MeSH descriptor within a specified branch level, ranging from the “Disease” main branch to subsequent levels such as the LV1 subbranch and beyond. We accomplished this by tallying the occurrences of each MeSH term across all publications in our database. After obtaining these raw counts, we moved to a critical phase of normalization. We normalized each count by the total number of articles within the database, thus converting raw frequencies into proportional measures. This adjustment allows the data to accurately reflect the prevalence of each descriptor within the context of the overall literature corpus.

For a more granular analysis of specific subbranches (i.e., LV1, LV2, etc), we refined our approach. We quantified the number of articles associated with each MeSH descriptor within the subbranch of interest. This time, however, the normalization process took into account the total number of articles relevant to that particular subbranch, thus ensuring that our statistical insights were accurately contextualized within the scope of the subbranch’s literature.

To ascertain the relative significance of specific MeSH descriptors within our databases, we denoted the frequency of each MeSH term within a database (DB) as \( f_{DB}(MeSH) \). This measure allows us to conduct a comparative analysis to determine the prominence of each descriptor in the COVID-19 database relative to the CVD database. We measure the disparity in usage frequency of a MeSH term between the 2 databases by calculating the difference, expressed as:

\[
\Delta f(MeSH) = f_{COVID}(MeSH) - f_{CVD}(MeSH)
\]

(1)

This difference, \( \Delta f(MeSH) \), provides an indication of whether a MeSH descriptor’s presence is more pronounced (up-regulated) or less pronounced (down-regulated) in the COVID-19 database as compared to the CVD database. A positive difference signifies a MeSH term’s greater relevance to the COVID-19 corpus, while a negative value indicates lesser importance.

However, the difference in frequencies can be misleading if the absolute values are too large or small. This difference might not accurately represent the term’s practical significance. To address this, we also calculated the ratio, \( R(MeSH) \), defined as:

\[
R(MeSH) = \frac{f_{COVID}(MeSH)}{f_{CVD}(MeSH)}
\]

(2)

This ratio offers insight into the relative usage of each MeSH descriptor. A ratio near 1 suggests comparable usage in both databases, while ratios significantly greater or less than 1 imply a disparity in descriptor usage.

By integrating both the difference, \( \Delta f(MeSH) \), and the ratio, \( R(MeSH) \), of frequencies, we achieved a more nuanced understanding of the role and emphasis of MeSH terms in the COVID-19 database in contrast with the CVD database. This dual-parameter approach allows for a more detailed and representative interpretation of the importance of specific MeSH descriptors in relation to the topics under investigation, such as hypertension.

Statistical Analysis

To pinpoint the most significant MeSH topics within the context of the 3 most prominent CVDs in relation to COVID-19, we employed a statistical approach. We conducted a chi-square test to determine if the frequency distribution of specific MeSH terms significantly differed across the COVID-19 and CVD datasets.
of independence. This statistical test was employed to assess whether the occurrence of specific MeSH terms shows a significant difference when comparing the COVID-19 database with the CVD database.

The chi-square test is particularly suited for this analysis as it helps determine if there is a significant association between the type of database (COVID-19 or CVD) and the frequency of particular MeSH terms. A significant result from this test implies that the likelihood of a MeSH term’s occurrence is dependent on the database, indicating a specific relevance to either COVID-19 or CVD-related articles.

Such a methodological approach allows us to identify and highlight those MeSH terms that are disproportionately represented in one database compared to the other, thereby providing insights into the intersection of COVID-19 with prominent cardiovascular conditions. This analysis not only enhances our understanding of disease dynamics but also potentially guides future research directions in these intersecting medical areas.

### Results

Our analysis begins with a thorough examination of the “Disease” main branch in the MeSH tree. Specifically, our interest lies in the corresponding MeSH descriptors found within the LV1 subbranches. These LV1 subbranches are particularly notable as they encompass the primary disease descriptors, which are the fundamental classifications for various diseases.

In Figure 3, we show the frequency with which primary disease descriptors are used in COVID-19 articles, \( f_{\text{COVID}}(\text{MeSH}) \). Additionally, we analyze articles that have been assigned the MeSH term “Comorbidity,” focusing exclusively on the frequency of disease descriptors within this subset. We compute the relative importance as defined in equation (1) to assess the importance of each LV1 disease descriptor in the context of comorbidity-related articles.

Overall, we find that all COVID-19 articles are labeled with the LV1 disease descriptor “Infections.” The second most common LV1 disease descriptor is “Respiratory Diseases,” which appears in 26.7% of all articles. The third descriptor, “Pathological Conditions, Signs and Symptoms,” was found in about 20% of all COVID-19 articles. The MeSH term “Cardiovascular Diseases” is the fourth most used descriptor, found in 7% of all articles. The top ten LV1 disease descriptors found in COVID-19 articles are shown in Figure 3A. In contrast, the results in Figure 3B illustrate the relative importance of the disease descriptors in a subset of COVID-19 articles related to comorbidities, considering the baseline frequency shown in Figure 3A. Therefore, the results in Figure 3B evaluate the importance of each LV1 disease descriptor specifically for these selected articles. As can be seen in Figure 3B, “Cardiovascular Diseases” has the highest relative importance among LV1 disease descriptors in COVID-19 articles examining comorbidity. This MeSH term has a 17.4% higher frequency of occurrence among COVID-19 articles related to comorbidities. It is also interesting to note that the MeSH term “Nutritional and Metabolic Diseases” ranks second.

In continuation, we focus on the first- and second-ranked MeSH descriptors in Figure 3B. To describe the disease terms in more detail, we repeat the analysis at the second level of the MeSH disease tree. Again, we separately calculated the proportion of items with a given LV2 disease descriptor and the relative importance of these descriptors within COVID-19 articles related to comorbidities. The results are shown in Figure 4.

From the results shown in Figure 4A, we see that “vascular disease” and “Heart Disease,” which belong to “Cardiovascular Diseases,” are among the 10 most frequently used LV2 disease descriptors. For COVID-19 articles related to comorbidity, both “Vascular Disease” and “Heart Disease” gain prominence (Figure 4B). The LV1 subbranch “Cardiovascular Diseases” is divided into 5 MeSH descriptors at the second level (Cardiovascular Abnormalities, Cardiovascular Infections, Heart Diseases, Pregnancy Complications, Cardiovascular and Vascular Diseases). In contrast, the disease branch “Nutritional and Metabolic Diseases” is divided into 2 descriptors at the second level (Metabolic Diseases, Nutritional Disorders). This should be considered as it could lead to a bias in the frequency of occurrence of a descriptor caused by the number of terms available in each subbranch. However, since we are interested in their relative importance, we can circumvent these biases and reveal the distributed or concentrated importance of the descriptors. Therefore, we continue our analysis at the third level of the MeSH tree of diseases. Since we found that at the second level of the MeSH tree, the LV1 subbranch “Cardiovascular Diseases” and “Nutritional and Metabolic Diseases” have the highest relative importance, we continue our investigation in this direction. The results are shown in Figure 5.

Figure 5A reveals that “Disorders of Glucose Metabolism” top the list as the most frequently mentioned LV3 MeSH term within the COVID-19 data set, followed by “Hypertension” and “Disease Attributes.” This figure provides an overarching view of the commonality of these terms across all research articles.

Figure 5B delves into the LV3 MeSH descriptors that stem from the LV2 subbranch of “Metabolic Diseases”. The data clearly indicate that disorders of glucose and lipid metabolism are the most recurrent topics within the LV3 subbranch, underscoring their significance in the discourse on metabolic diseases.

Figure 5C illustrates that within the realm of CVDs, “Embolism and Thrombosis” emerges as the most prevalent LV3 MeSH descriptor utilized in the literature, followed by “Hypertension” and “Heart Failure,” among others, in descending order of frequency.

By comparing Figure 5B and 5C, we observe a less diverse distribution of the embedded MeSH terms. The LV3 descriptors within the CVD subbranch are more specifically clustered, pointing to a narrower focus within CVD research in relation to COVID-19, as opposed to the broader range of topics covered under metabolic diseases.

Figure 5D presents a detailed ranking of LV3 MeSH descriptors within the CVD domain as they appear in the context of comorbidity research. Figure 5D specifically highlights which cardiovascular conditions are most frequently discussed in
conjunction with other health issues, shedding light on the patterns of comorbidity that are prevalent in the current body of literature. It allows researchers to identify which cardiovascular disorders are most considered in studies that address the complexities of patients presenting with multiple concurrent health challenges. The data presented in Figure 5D identifies “Hypertension” as the most used LV3 MeSH descriptor within articles that discuss comorbidities, with “Heart Failure” following in frequency. Based on these observations from Figure 5, the subsequent analysis concentrates on 3 critical LV3 MeSH terms: “Hypertension,” “Heart Failure,” and “Embolism and Thrombosis.”

In our analysis of the COVID-19 and CVD databases, we use the frequency of LV3 MeSH descriptors to represent the focus of research. We started by examining “Embolism and Thrombosis,” a common CVD descriptor (Figure 5C). Our results (Figure 6A) indicate that “Embolism and Thrombosis” is most frequently associated with “Disorders of Glucose Metabolism” in the COVID-19 database. “Disorders of Iron Metabolism” (with an increase of $\Delta f_{[MeSH]}=0.16\%$ and a ratio of $R_{[MeSH]}=5.64$) and “Disorders of Acid-Base Balance” are also significant but less frequent. “Disorders of Iron Metabolism” have seen the largest increase, ranking it at the top in the COVID-19 database. “Disorders of Glucose Metabolism” follow (with an increase of $\Delta f_{[MeSH]}=1.14\%$ and a ratio of $R_{[MeSH]}=2.46$), and “Disorders of Acid-Base Balance” come in third (with an increase of $\Delta f_{[MeSH]}=0.06\%$ and a ratio of $R_{[MeSH]}=1.47$).

**Figure 3.** Comparative analysis of level 1 (LV1) subbranch disease descriptor frequencies in COVID-19–related articles. Panel A presents the distribution within the COVID-19 data set, while Panel B focuses on the subset of COVID-19 articles tagged with the “Comorbidity” MeSH (Medical Subject Headings) term. Each bar’s color corresponds to a specific disease descriptor and maintains consistency throughout the manuscript.

**Figure 4.** Most important level 2 (LV2) disease descriptor. Results are computed for (A) the entire COVID-19 data set and (B) for the subset of COVID-19 articles related to comorbidities. Each bar’s color corresponds to a specific disease descriptor, as defined in Figure 3. MeSH: Medical Subject Headings.
Figure 5. Ten most important level 3 (LV3) disease descriptors for COVID-19–related articles. Results are computed for (A) all LV3 disease descriptors, (B) only for LV3 disease descriptors originating from the “Nutritional and Metabolic Diseases,” (C) only for the LV3 “Cardiovascular Diseases” branch, and (D) only for the LV3 “Cardiovascular Diseases” branch obtained for the COVID-19 sub-set of articles considering comorbidities. Each bar’s color corresponds to a specific disease descriptor, as defined in Figure 3. MeSH: Medical Subject Headings.
Figure 6. A comprehensive comparison of the absolute and relative changes in the frequency of level 3 (LV3) MeSH (Medical Subject Headings) descriptors in the COVID-19 database relative to the cardiovascular disease (CVD) database, focusing on 3 specific cardiovascular diseases: Embolism and Thrombosis (panels A and B), Hypertension (panels C and D), and Heart Failure (panels E and F). The figure displays the top 15 MeSH descriptors for each condition. Notably, the bars corresponding to “Iron Metabolism Disorders,” “Glucose Metabolic Disorders,” “Acid-Base Imbalance,” and “Disorders of Lipid Metabolism” are distinctly color-coded in black, red, green, and orange, respectively, allowing for easy identification and comparison of these key terms across different cardiovascular conditions.

For high blood pressure (“Hypertension”), the MeSH term “Disorders of Glucose Metabolism” is most significant in frequency difference, followed by “Acid-Base Imbalance” and “Iron Metabolism Disorders” (Figure 6C). Interestingly, “Iron Metabolism Disorders” show a smaller overall frequency difference ($\Delta f[\text{MeSH}]=0.11\%$) but a higher ratio ($R[\text{MeSH}]=5.90$), indicating they are used more frequently in COVID-19 research compared to CVD research. “Disorders of Glucose Metabolism” take the second spot ($\Delta f[\text{MeSH}]=26.47\%$ and $R[\text{MeSH}]=3.66$), and “Acid-Base Imbalance” is third ($\Delta f[\text{MeSH}]=0.21\%$ and $R[\text{MeSH}]=2.10$) based on their relative frequencies.

In the third part of our analysis, we focused on heart failure, with the findings illustrated in Figure 6E and 6F. Among the various MeSH terms, “Glucose Metabolic Disorders” emerged as the second most frequent term in the comparison between the COVID-19 and CVD databases. While “Iron Metabolic
Disorders” and “Acid-Base Imbalance” are also relevant, they are positioned at 17th, explaining their absence from the figure due to their lower frequency. Notably, “Iron Metabolism Disorders” feature more prominently in the COVID-19 database than in the CVD database, ranking tenth in frequency difference. Significantly, as Figure 6F reveals, “Iron Metabolism Disorders” rank second in relative importance among all LV3 MeSH descriptors in the COVID-19 database, compared to the CVD database. “Disorders of Lipid Metabolism” also show considerable relevance, ranking third, whereas “Disorders of Glucose Metabolism” are positioned seventh. Despite its lower frequency, “Acid-Base Imbalance” maintains a high relative importance, coming in at 17th. These results underscore the shifted focus in medical research on specific metabolic disorders in the context of COVID-19, particularly in relation to heart failure.

To build upon these findings, we applied the chi-square test to validate whether the observed differences in LV3 MeSH descriptor frequencies between COVID-19 and CVD databases are statistically significant. This test helped us determine if the occurrences of 4 specific MeSH terms—“Disorders of Glucose Metabolism,” “Iron Metabolism Disorders,” “Acid-Base Imbalance,” and “Disorders of Lipid Metabolism”—in the COVID-19 database are significantly different from their occurrences in the CVD database. We also employed multiple P values to strengthen our assessment of significance. The findings are detailed in Table 1.

**Table 1. Statistical significance of selected MeSH (Medical Subject Headings) terms in 3 subsets of COVID-19 articles related to cardiovascular diseases (CVDs).**

<table>
<thead>
<tr>
<th>LV3 MeSH terms</th>
<th>Embolism and thrombosis, P value</th>
<th>Hypertension, P value</th>
<th>Heart failure, P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose Metabolism Disorders</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Iron Metabolism Disorders</td>
<td>&lt;.01</td>
<td>.525</td>
<td>.116</td>
</tr>
<tr>
<td>Acid-Base Imbalance</td>
<td>.103</td>
<td>.42</td>
<td>.585</td>
</tr>
<tr>
<td>Lipid Metabolism Disorders</td>
<td>.94</td>
<td>.53</td>
<td>.05</td>
</tr>
</tbody>
</table>

aThe P values signify whether the appearance of a MeSH term in the COVID-19 database is significantly different compared to the appearance in the CVD database.
bLV3: level 3.

In the context of “Embolism and Thrombosis,” our analysis reveals that the frequencies of both glucose and iron metabolism disorders show a statistically significant difference when comparing the COVID-19 and CVD databases across all 3 subdata sets. For “Hypertension,” the scenario is slightly different. Here, the incidence of glucose metabolism disorders stands out as the only descriptor with a significant difference in frequency between the COVID-19 and CVD databases. Lastly, regarding “Heart Failure,” we again note a significant difference for glucose metabolism disorders, although with a P value of <.05. This pattern highlights a specific focus or heightened research interest in glucose metabolism disorders within the context of COVID-19, particularly when comorbid with CVDs such as embolism, thrombosis, hypertension, and heart failure.

**Discussion**

**Principal Results**

The aim of this study was to analyze all available peer-reviewed articles from the PubMed database to identify the most relevant topics regarding the relationship among COVID-19, CVDs, and comorbidity. For this purpose, we used the MeSH term descriptors, which are the most important topics covered in an article in a standardized form. In COVID-19–related research that considers comorbidity, we found the most relevant MeSH descriptors are CVDs and nutritional and metabolic diseases. Since both terms are quite broad, we continued our analysis one branch deeper in the MeSH tree and found that the corresponding significant topics are related to metabolic disorders, vascular diseases, and heart diseases. Advancing one level deeper in the MeSH tree, we investigated the meaning of more specific terms related to COVID-19 and metabolic disease. We determined that the most significant CVDs related to comorbidity and COVID-19 are embolism and thrombosis, hypertension, and heart failure. Given the prominence of metabolic disorders in our analysis, we also explored which specific metabolic disorders were most significant and found that glucose metabolism disorders were the most notable. However, we also noted a significantly increased frequency of the term iron metabolism disorders in COVID-19 articles related to embolism and thrombosis compared to CVD articles related to embolism and thrombosis.

**Limitations**

Using the methodology presented here, we were able to identify the most important issues relevant to comorbidities and COVID-19. Although the methodology can be applied to any major topic and its corresponding subtopic, it has some limitations. The main limitation is its inability to find relationships between themes. This was addressed by selecting relevant subtopics through iteratively evaluating the results at each level of the MeSH tree. However, in future studies, we intend to incorporate a knowledge graph-based approach by mapping relationships between topics. This would in turn allow us to consider not only the frequency of the occurrence of a topic but also to evaluate the co-occurrence of topics. Consequently, this would allow us to automatically find highly related pairs of topics and eventually create a more detailed and complex description of the item database under consideration.
Comparison With Prior Work

In relation to COVID-19, individuals with certain comorbidities have been shown to have a higher likelihood of developing a severe form of this disease and have a higher mortality rate. COVID-19 has been associated with an increased prevalence of CVD, suggesting that CVD may be a risk factor for the disease [50]. According to mortality data from China’s National Health Commission, 17% and 35% of individuals with COVID-19 had a history of coronary heart disease and hypertension, respectively [51]. Li et al [52] showed that the presence of cardio-cerebrovascular disease, diabetes, and hypertension increased the risk of severe COVID-19 by threefold, twofold, and twofold, respectively. A larger study from the Chinese Center for Disease Control and Prevention, which examined the clinical outcomes of 44,672 confirmed COVID-19 cases, found that the case fatality rate was 2.3% in the entire cohort, but was significantly higher (6%, 7.3%, and 10.5%, respectively) in individuals with hypertension, diabetes, and CVD [53]. Several smaller cohort studies have also presented similar reports, suggesting a higher risk of an adverse episode in patients with COVID-19 with underlying CVD [54-56]. Cardiac injury (characterized by elevated troponin levels), myocarditis, and acute respiratory distress syndrome have been reported as strong, independent risk factors associated with mortality in patients with COVID-19 [57]. According to the Pneumonitis Diagnosis and Treatment Program for Novel Coronavirus Infections, the likelihood of COVID-19 infection is higher in older people (>60 years) with pre-existing conditions, especially in patients with hypertension, coronary heart disease, or diabetes [51]. Thus, advanced age, male gender, and the presence of preexisting conditions are the main risk factors for COVID-19 mortality [57].

Given the increasing evidence of iron status’ importance for immunity, it is not surprising that biomarkers of iron metabolism have been investigated in several studies on patients with COVID-19 [58]. COVID-19 is also characterized by a cytokine storm, leading to increased production of hepcidin, the primary hormone regulating iron metabolism, in response to heightened proinflammatory cytokines [59]. Patients with low serum iron status were likely to suffer from severe conditions and multiple organ damage in COVID-19 [60]. In addition, both iron deficiency (ID) and iron overload are commonly observed in a variety of CVDs and contribute to the onset and progression of these diseases. One of the devastating consequences of iron overload is the induction of ferroptosis, a newly defined form of regulated cell death that severely impairs cardiac function through ferroptotic cell death in cardiomyocytes [60]. Our results show that the term iron metabolism disorder occurs significantly more frequently in COVID-19 articles related to heart failure than in CVD articles on the same topic. Interestingly, ID is frequently observed in patients with heart failure [61-63]. Furthermore, ID correlates with an increased incidence of right ventricular failure in patients with acute HF [64,65]. ID also contributes to impaired functioning of the respiratory chain complexes (complex I to V), leading to altered myocardial metabolism, ROS formation, and ultimately advanced HF. Impaired mitochondrial function is one of the underlying mechanisms of ID-induced HF [66,67].

Clinical Implications of Our Findings

Our research has highlighted the critical intersection between COVID-19 and severe cardiovascular conditions, notably embolism and thrombosis. The urgency of identifying and managing these conditions is of paramount importance, as they present immediate life-threatening risks and their symptoms often overlap with those of COVID-19, especially pulmonary thromboembolism [68]. Our findings underscore the vital importance of vigilant monitoring for individuals affected by COVID-19 to prevent these severe outcomes.

A primary tool in this monitoring process is the serial measurement of D-dimer levels, which has been shown to strongly correlate with an increased risk of disease progression, critical illness, and mortality. D-dimer levels also serve as a reliable predictor of venous thromboembolism when measured at admission, and levels at discharge are associated with a higher 1-year mortality risk [69]. Current guidelines recommend thromboprophylaxis for all hospitalized patients with COVID-19, except those with an increased risk of bleeding [68]. While further research is necessary to determine the optimal anticoagulation dosage, standard doses of LMWH are generally recommended for most patients, with intermediate doses for those who are critically ill or obese [70]. Routine screening for deep vein thrombosis with Doppler ultrasonography is not currently advised for thromboembolism screening, as rapidly increasing D-dimer levels and worsening oxygenation have been found to be more successful [71].

We have also uncovered a significant correlation between COVID-19 and glucose metabolism disorders. Increasing evidence suggests a bidirectional relationship between diabetes and SARS-CoV-2 infection. This indicates that patients with diabetes are at a higher risk of developing a severe form of COVID-19, while individuals with COVID-19 are more likely to develop metabolic disorders. Shared pathogenic mechanisms, such as general inflammation, a pro-thrombotic state, and atherosclerosis, likely contribute to this association [72].

Analysis of the GTEx database revealed higher ACE2 expression in the pancreas than in the lungs. Liu et al. analyzed pancreatic injury following SARS-CoV-2 infection and found that such injuries predominantly occurred in patients with severe COVID-19 [73]. Therefore, special attention is warranted for patients with metabolic disorders, including priority for vaccination and rigorous monitoring in the event of infection, with a low threshold for intensifying care. Preventive measures for detecting metabolic disorders should be implemented in individuals after a severe SARS-CoV-2 infection. This includes monitoring blood glucose levels, lipids, and biochemical markers for pancreatic injury.

Additionally, our results underscore the importance of iron metabolism, a factor currently underrepresented in clinical practice, underscoring the need for further trials to integrate it into care for patients with COVID-19. Research indicates that ferritin levels can be used to estimate disease severity, providing useful cutoff values. These could complement other initial screening methods in predicting the necessary level and intensity of patient care [74]. There is also an underexplored therapeutic potential in manipulating iron levels, either by using chelators
like deferoxamine to lower them or through iron supplementation to raise them in patients with inappropriate values. Before this approach can be widely adopted in practice, further research is essential to determine the optimal levels. Reducing iron in patients with highly active hepcidin due to inflammation could impede recovery [75]. Nonetheless, the importance of iron metabolism extends beyond coagulation disorders to metabolic disorders, with iron overload contributing to the development of these diseases [76,77].

In summary, the insights from our study have critical implications for clinical practice. By identifying key biomarkers and conditions associated with severe COVID-19 outcomes, we provide a foundation for improving patient monitoring, treatment strategies, and, ultimately, patient outcomes during the pandemic. Our findings urge health care professionals to incorporate these insights into their clinical practice, promoting a proactive and informed approach to managing COVID-19 and its cardiovascular complications.

Conclusions

Our study represents a crucial step toward understanding the complex interplay between COVID-19, CVD, and metabolic disorders, highlighting in particular the role of embolism, thrombosis, and iron metabolism disorders. The method we adopted, using MeSH term descriptors to dissect the different levels of related topics, has furnished a comprehensive overview of the main comorbidities influencing COVID-19 outcomes. Importantly, this approach can be adapted and applied to other important health topics and their subcategories, despite its current limitation in directly mapping topic relationships. Future research efforts should aim to incorporate a knowledge graph–based methodology, enabling a more detailed analysis of topic co-occurrences and their relationships. Such advancements are essential for deciphering complex disease dynamics, particularly in the context of emerging infectious diseases such as COVID-19.

The knowledge gained from this study is invaluable for the development of more effective clinical practices and public health strategies. By identifying key comorbidities and their impact on COVID-19, we are better positioned to tailor treatments and interventions for patients affected by these conditions. Additionally, understanding the role of specific metabolic disorders, such as those affecting glucose and iron metabolism, opens up potential therapeutic targets and preventive measures. In managing the current pandemic and preparing for future viral outbreaks, the findings from this study are crucial in guiding medical advancements, improving patient outcomes, and increasing the resilience of the health care system. This work not only contributes to our immediate fight against COVID-19 but also creates a foundation for more informed and effective responses to similar health crises in the future.

Acknowledgments

This research was supported by the Slovenian Research Agency (Javna agencija za raziskovalno dejavnost RS; grant P1-0403, J1-2457, N1-0232, J3-3077, J5-3114, P1-0055, and I0-0029).

Authors’ Contributions

MM, RM, and VG were involved in the design of this study. RM, LT, TT, and MM wrote the introduction. Datamining and data analysis were performed by RM. Data visualization was performed by RM and VG. All authors critically revised the manuscript. The final manuscript has been read and approved by all authors.

Conflicts of Interest

None declared.

References


Abbreviations

ACE2: angiotensin-converting enzyme 2  
CVD: cardiovascular disease  
DB: database  
ID: iron deficiency  
LV1: level 1  
LV2: level 2  
LV3: level 3  
MeSH: Medical Subject Headings  
NLM: National Library of Medicine  
PMID: PubMed identifier  
RAAS: renin-angiotensin-aldosterone system
Factors Associated With Worsened Mental Health of Health Care Workers in Canada During the COVID-19 Pandemic: Cross-Sectional Survey Study

Ali AbdulHussein¹, BEng, MEng, MBA; Zahid Ahmad Butt², PhD; Stanko Dimitrov¹, PhD; Brian Cozzarin¹, PhD

¹Department of Management Science, Faculty of Engineering, University of Waterloo, Waterloo, ON, Canada
²School of Public Health Sciences, Faculty of Health, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Ali AbdulHussein, BEng, MEng, MBA
Department of Management Science
Faculty of Engineering
University of Waterloo
200 University Ave W
Waterloo, ON, N2L 3G1
Canada
Phone: 1 (519) 888 4567
Fax: 1 (519) 888 4567
Email: a9abdulh@uwwaterloo.ca

Abstract

Background: Health care workers (HCWs) in Canada have endured difficult conditions during the COVID-19 pandemic. Many worked long hours while attending to patients in a contagious environment. This introduced an additional burden that may have contributed to worsened mental health conditions.

Objective: In this study, we examine the factors associated with worsened mental health conditions of HCWs as compared to before the start of the pandemic.

Methods: We use data from a survey of HCWs by Statistics Canada. A regression model is used to estimate the odds ratios (ORs) of worsened mental health after the start of the pandemic. The estimated odds ratio (OR) is associated with different independent variables that include demographics (age, sex, immigration status, and geographic area), occupational factors (work status, occupational group, and exposure category), and different access levels to personal protective equipment (PPE).

Results: Of 18,139 eligible participants surveyed, 13,990 (77.1%) provided valid responses. We found that HCWs younger than 35 years old were more likely (OR 1.14, 95% CI 1.03-1.27; P=.01) to exhibit worsened mental health as compared to the reference group (35-44 years old). As for sex, male HCWs were less likely (OR 0.76, 95% CI 0.67-0.86; P<.001) to exhibit worsened mental health as compared to female HCWs. Immigrant HCWs were also less likely (OR 0.57, 95% CI 0.51-0.64; P<.001) to exhibit worsened mental health as compared to nonimmigrant HCWs. Further, HCWs working in Alberta had the highest likelihood of exhibiting worsened mental health as compared to HCWs working elsewhere (Atlantic provinces, Quebec, Manitoba, Saskatchewan, Ontario, British Columbia, and Northern Territories). Frontline workers were more likely (OR 1.26, 95% CI 1.16-1.38; P<.001) to exhibit worsened mental health than nonfrontline HCWs. Part-time HCWs were less likely (OR 0.85, 95% CI 0.76-0.93; P<.001) to exhibit worsened mental health than full-time HCWs. HCWs who reported encountering COVID-19 cases were more likely (OR 1.55, 95% CI 1.41-1.70; P<.001) to exhibit worsened mental health than those who reported no contact with the disease. As for PPE, HCWs who never had access to respirators, eye protection, and face shields are more likely to exhibit worsened mental health by 1.31 (95% CI 1.07-1.62; P<.001), 1.51 (95% CI 1.17-1.96; P<.001), and 1.41 (95% CI 1.05-1.92; P=.02) than those who always had access to the same PPE, respectively.

Conclusions: Different HCW groups experienced the pandemic differently based on their demographical and occupational backgrounds as well as access to PPE. Such findings are important to stakeholders involved in the planning of personalized support programs and aid mental health mitigation in future crises. Certain groups require more attention.

doi:10.2196/50064

KEYWORDS
health care workers; COVID-19; mental health; demographic factors; occupational factors; access to PPE; pandemic; health care system; psychological trauma; psychological; trauma; respirators; eye protection; face shields; support

Introduction
On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic. The pandemic resulted in devastating health impacts on populations and a crisis within the health care system [1]. This health care system was tasked to handle the unprecedented inflow of patients. Functions within the system that were impacted include emergency departments, intensive care units, physician services, and long-term care units [2]. Health care workers (HCWs) of different occupational groups battled the pandemic. Overwhelmed hospitals in Canada canceled less urgent surgeries by up to 80% by June 2020 [3]. These patterns shifted the workload on HCWs and had an impact on the overall health care system. These conditions not only demanded more hospital capacity but also put an overwhelming strain on HCWs [4]. In addition to the operational pressure, HCWs also suffered from the lack of personal protection equipment (PPE), especially at the beginning of the pandemic [5]. HCWs were at the frontline in battling this pandemic. This battle has put pressure on their mental health conditions [6-8].

In this paper, we assess the various factors associated with worsened HCWs’ mental health conditions as compared to before the start of the pandemic. Although experiencing mental health conditions may be a daily occurrence for some HCWs, the duration and severity during the pandemic were different. HCWs were also at a higher risk of infection, adding to the risk of further mental health conditions [9]. In the past, HCWs have experienced mental health problems during other outbreaks, including the Middle East respiratory syndrome and the severe acute respiratory syndrome [10]. Such conditions have been studied in the literature [11]. In this study, we use a recent data set by Statistics Canada from a national cross-sectional survey that was conducted in the fall of 2020 to assess the impact of COVID-19 on HCWs. Unlike work in the literature, we comprehensively assess the impact of demographic and occupational factors as well as the availability of PPE on HCWs’ mental health conditions [9,12,13].

In the global literature, authors assessed the mental health conditions of HCWs during such outbreaks globally [14-16]. The impact of demographic, social, and occupational factors was reportedly linked to various mental health conditions. Researchers also assessed the prevalence of stress, anxiety, and other psychological well-being indicators of HCWs in Oman during the pandemic [12]. The focus was on young female HCWs who encountered confirmed or suspected COVID-19 cases during their work. Another study in Turkey examined the relationship between the perceived risk of infection and the mental health conditions of HCWs during the COVID-19 pandemic [13]. In addition, a study in the Chinese province of Hubei was conducted early in the pandemic to assess the psychological impact of the pandemic on the frontline medical staff [9]. The study measured the association of factors including professional group, age, and sex factors with work stress.

Our study is among the first to highlight the association of diverse demographic and occupational factors with the mental health condition of HCWs during the pandemic in Canada. We also considered the role of access to PPE on HCWs’ mental health conditions. Unlike existing literature, we studied the individual impact of each of the following factors on the mental health of HCWs while holding the rest constant. Demographic factors include age group, sex, province of the workplace, and immigration status. Occupational factors include work status, frontline category, and exposure to confirmed or suspected COVID-19 cases. Finally, we also considered access to several PPE. The findings of the study will be of prime importance to key stakeholders, including mental health support program planners, health care policy makers, HCWs themselves, and researchers in the area. The goal of this paper is to understand which factors were associated with worsened mental health conditions in HCWs after the start of the pandemic.

Methods

Data Sources, Study Procedure, and Participants
We used a data set from a recent cross-sectional survey by Statistics Canada on the impacts of COVID-19 on HCWs. Unlike other traditional Statistics Canada surveys, a random selection of participants was not used. Instead, Statistics Canada sent an email invitation to HCWs across Canada. Then, a snowball sampling procedure was used. The invitation included a link to a web-based survey that was available through Statistics Canada’s web page. Accordingly, 18,139 responses were collected between November 24 and December 13, 2020, across 7 provincial and territorial regions in Canada. No data were collected for this survey beyond these dates. Only responses by HCWs were included in the data set. The responses were completely anonymized by Statistics Canada.

The questionnaire asked HCWs for information related to the job environment, demographics, geography, and information on access to PPE as background information. Adaptive questioning was used. Our study was limited to these 3 factors categories only: demographics, occupational, and access to PPE based on the available data from Statistics Canada. In assessing mental health well-being, the survey asked HCWs: compared to before the COVID-19 pandemic, how would you say your mental health is now? HCWs self-reported their perceived mental health on a 5-point Likert scale: much better now, somewhat better now, about the same, somewhat worse now, much worse now.

Statistical Analysis Strategy
Our study uniquely studied the association of various interesting factors, as depicted in Figure 1, with HCW mental health conditions as compared to before the start of the pandemic. To consider each factor separately, we used a multivariate ordinal logistic regression [17,18]. The model is defined in Multimedia Appendix 1. The dependent variable is the state of mental health
of the respondents. We reduced the mental health state categories from the 5 mentioned above to 3 (improved, same, or worsened) to yield statistically significant model estimates. The independent variables include demographics such as age, sex, immigration status, and the province of the workplace. Occupational variables include work status, occupational group, and exposure to COVID-19 cases. The model also estimated the association of HCWs’ access to a variety of PPE, such as respirators, eye protection, and face shields. Access to PPE is also reported by the HCW on a 5-point Likert scale, which we reduced to 3 categories (always available, sometimes available, or never available) to yield statistically significant model estimates. All responses are self-reported. Analyses were conducted in RStudio (version 1.4.1717; Posit, PBC).

**Figure 1.** Factors associated with the mental health of HCWs. HCW: health care worker; PPE: personal protective equipment.

### Ethical Considerations
Data collected by Statistics Canada were reviewed based on the following principles: privacy, security, transparency, accountability, trust, sustainability, data quality, and fairness, as well as well vetted to be harmless to participants or the public. All ethical approvals were considered by Statistics Canada. No further ethical approval to use these data was required by the authors and the survey is made available to the public. This study is exempt from further ethical approval. Furthermore, the authors also did not have access to information related to the development and testing of the survey. The authors are also not aware of any compensation provided to survey participants.

### Results

#### Participant Profile
From the full data set of 18,139 responses, we considered nonduplicate responses that provided valid answers to all questions of interest to this study (N=13,990). The remaining responses contained missing or invalid responses. For the considered population, Table 1 outlines the number of observations and percentages for different demographic factors such as age group, sex, and immigration status. The number of observations and percentages were also calculated for occupational factors including the province of the primary workplace, work status (full-time vs part-time), frontline work status, occupational group, and exposure to confirmed or suspected COVID-19 cases. Additionally, the last 3 rows of Table 1 outline PPE access levels for 3 different PPE: face shields, eye protection, and respirators. Overall 4261 (31%) HCWs were younger than 35 years. The majority were female participants (n=12,682, 91%). In total, 12,510 (89%) HCWs were nonimmigrant. HCW workers were distributed across 7 regions, with the largest group of 6626 (47%) working in Ontario. The majority (n=10,152, 73%) of HCWs worked full-time and 5511 (39%) were nonfrontline allied health professionals. A total of 8188 (59%) HCWs did not report exposures to confirmed or suspected COVID-19 cases. As for access to PPE, 10,758 (77%) HCWs always had access to face shields, 10,118 (72%) always had access to eye protection, and 6898 (49%) always had access to respirators.
Table 1. Participant demographic, occupational, and PPE\textsuperscript{a} access characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>4261 (30.5)</td>
</tr>
<tr>
<td>35-44</td>
<td>3962 (28.3)</td>
</tr>
<tr>
<td>45-54</td>
<td>3415 (24.4)</td>
</tr>
<tr>
<td>≥55</td>
<td>2352 (16.8)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1308 (9.3)</td>
</tr>
<tr>
<td>Female</td>
<td>12,682 (90.7)</td>
</tr>
<tr>
<td><strong>Immigration status</strong></td>
<td></td>
</tr>
<tr>
<td>Nonimmigrant</td>
<td>12,510 (89.4)</td>
</tr>
<tr>
<td>Immigrant</td>
<td>1480 (10.6)</td>
</tr>
<tr>
<td><strong>Province of workplace</strong></td>
<td></td>
</tr>
<tr>
<td>Atlantic provinces</td>
<td>1972 (14.1)</td>
</tr>
<tr>
<td>Quebec</td>
<td>832 (5.9)</td>
</tr>
<tr>
<td>Ontario</td>
<td>6626 (47.4)</td>
</tr>
<tr>
<td>Manitoba and Saskatchewan</td>
<td>1308 (9.3)</td>
</tr>
<tr>
<td>Alberta</td>
<td>2018 (14.4)</td>
</tr>
<tr>
<td>British Columbia</td>
<td>1194 (8.5)</td>
</tr>
<tr>
<td>Northern territories</td>
<td>40 (0.3)</td>
</tr>
<tr>
<td><strong>Work status</strong></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>10,152 (72.6)</td>
</tr>
<tr>
<td>Part-time</td>
<td>3838 (27.4)</td>
</tr>
<tr>
<td><strong>Frontline occupation</strong></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>397 (2.8)</td>
</tr>
<tr>
<td>Nurse</td>
<td>4689 (33.5)</td>
</tr>
<tr>
<td>Emergency medical personnel</td>
<td>223 (1.6)</td>
</tr>
<tr>
<td><strong>Nonfrontline occupation</strong></td>
<td></td>
</tr>
<tr>
<td>Personal support worker</td>
<td>454 (3.2)</td>
</tr>
<tr>
<td>Allied health professional</td>
<td>5511 (39.4)</td>
</tr>
<tr>
<td>Laboratory worker</td>
<td>1267 (9.1)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>169 (1.2)</td>
</tr>
<tr>
<td>Dental professional</td>
<td>1280 (9.1)</td>
</tr>
<tr>
<td><strong>Exposure to confirmed or suspected cases</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5802 (41.5)</td>
</tr>
<tr>
<td>No</td>
<td>8188 (58.5)</td>
</tr>
<tr>
<td><strong>Access to PPE</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Face shields</strong></td>
<td></td>
</tr>
<tr>
<td>Always or usually available</td>
<td>10,758 (76.9)</td>
</tr>
<tr>
<td>Sometimes available</td>
<td>1070 (7.6)</td>
</tr>
<tr>
<td>Never available</td>
<td>357 (2.6)</td>
</tr>
<tr>
<td>Skipped answer</td>
<td>1805 (12.9)</td>
</tr>
<tr>
<td><strong>Eye protection</strong></td>
<td></td>
</tr>
<tr>
<td>Characteristics</td>
<td>Value, n (%)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Always or usually available</td>
<td>10,118 (72.3)</td>
</tr>
<tr>
<td>Sometimes available</td>
<td>1132 (8.1)</td>
</tr>
<tr>
<td>Never available</td>
<td>556 (4.0)</td>
</tr>
<tr>
<td>Skipped answer</td>
<td>2184 (15.6)</td>
</tr>
</tbody>
</table>

**Respirators**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always or usually available</td>
<td>6898 (49.3)</td>
</tr>
<tr>
<td>Sometimes available</td>
<td>1369 (9.8)</td>
</tr>
<tr>
<td>Never available</td>
<td>676 (4.8)</td>
</tr>
<tr>
<td>Skipped answer</td>
<td>5047 (36.1)</td>
</tr>
</tbody>
</table>

**Analysis Results**

In the following, we present the results for estimating the associations between the various factors and the likelihood of worsened mental health conditions for HCWs.

**Demographic Factors**

Model estimates are expressed as odds ratios (ORs), as presented in Table 2. These ORs indicate the odds of worsened mental health conditions as compared to before the start of the pandemic. Based on the ORs in Table 2, HCWs who are younger than 35 years old were more likely (OR 1.14, 95% CI 1.03-1.27; \( P < .01 \)) to exhibit worsened mental health conditions than the reference group (35 to 44 years old). Furthermore, those aged 45-54 years and 55 years and older were less likely to exhibit worsened mental health conditions than the reference group (OR 0.71, 95% CI 0.64-0.78; \( P < .001 \); and OR 0.55, 95% CI 0.49-0.61; \( P < .001 \), respectively). Hence, the older the HCW, the lower the likelihood of worsened mental health conditions.

As for sex, male HCWs were less likely (OR 0.76, 95% CI 0.67-0.86; \( P < .001 \)) to exhibit worsened mental health conditions than their female counterparts. Immigrant HCWs were also less likely (OR 0.57, 95% CI 0.51-0.64; \( P < .001 \)) to exhibit worsened mental health conditions than nonimmigrants.

Geographically, HCWs working in Alberta have the highest likelihood of worsened mental health conditions. HCWs living in Alberta were most likely to exhibit worsened mental health conditions.
Table 2. Estimates for the ordinal regression model for various factors associated with the mental health conditions of HCWs\textsuperscript{a}.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>$\beta$</th>
<th>SE</th>
<th>OR\textsuperscript{b} (95% CI)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group (years; reference: 35 to 44 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 35</td>
<td>0.13</td>
<td>0.05</td>
<td>1.14 (1.03-1.27)</td>
<td>.01</td>
</tr>
<tr>
<td>45-54</td>
<td>−0.35</td>
<td>0.05</td>
<td>0.71 (0.64-0.78)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>55 and older</td>
<td>−0.60</td>
<td>0.06</td>
<td>0.55 (0.49-0.61)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Sex (reference: female)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>−0.28</td>
<td>0.06</td>
<td>0.76 (0.67-0.86)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Immigration status (reference: nonimmigrant)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immigrant</td>
<td>−0.56</td>
<td>0.06</td>
<td>0.57 (0.51-0.64)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Work location (reference: Alberta)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic provinces</td>
<td>−0.41</td>
<td>0.07</td>
<td>0.66 (0.57-0.77)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Quebec</td>
<td>−0.30</td>
<td>0.09</td>
<td>0.74 (0.61-0.89)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Manitoba and Saskatchewan</td>
<td>−0.09</td>
<td>0.08</td>
<td>0.91 (0.77-1.08)</td>
<td>.27</td>
</tr>
<tr>
<td>Ontario</td>
<td>−0.24</td>
<td>0.06</td>
<td>0.78 (0.70-0.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>British Columbia</td>
<td>−0.09</td>
<td>0.09</td>
<td>0.92 (0.77-1.09)</td>
<td>.31</td>
</tr>
<tr>
<td>North Territories</td>
<td>−0.56</td>
<td>0.34</td>
<td>0.57 (0.30-1.15)</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Work status (reference: full-time)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part-time</td>
<td>−0.16</td>
<td>0.04</td>
<td>0.85 (0.76-0.93)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Job setting (reference: nonfrontline)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontline</td>
<td>0.23</td>
<td>0.05</td>
<td>1.26 (1.16-1.38)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Contact with patients with COVID-19 (reference: no contact)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure</td>
<td>0.44</td>
<td>0.05</td>
<td>1.55 (1.41-1.70)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Access to respirators (reference: did not need)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>−0.03</td>
<td>0.05</td>
<td>0.97 (0.88-1.07)</td>
<td>.13</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.30</td>
<td>0.08</td>
<td>1.34 (1.14-1.59)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>0.27</td>
<td>0.10</td>
<td>1.31 (1.07-1.62)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Access to eye protection (reference: did not need)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>−0.03</td>
<td>0.06</td>
<td>0.97 (0.85-1.10)</td>
<td>.17</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.29</td>
<td>0.11</td>
<td>1.33 (1.08-1.65)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>0.41</td>
<td>0.13</td>
<td>1.51 (1.17-1.96)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Access to face protection (reference: did not need)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>0.26</td>
<td>0.07</td>
<td>1.30 (1.14-1.48)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.40</td>
<td>0.11</td>
<td>1.50 (1.21-1.86)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>0.35</td>
<td>0.15</td>
<td>1.41 (1.05-1.92)</td>
<td>.02</td>
</tr>
</tbody>
</table>

\textsuperscript{a}HCW: health care worker.
\textsuperscript{b}OR: odds ratio.

**Occupational Factors**

Frontline HCWs, such as physicians, nurses, and emergency medical personnel, were more likely (OR 1.26, 95% CI 1.16-1.38; $P<.001$) to exhibit worsened mental health conditions than nonfrontline workers (personal support workers, allied health professionals, laboratory workers, pharmacists, and professionals). Part-time HCWs, however, were less likely (OR 0.85, 95% CI 0.76-0.93; $P<.001$) to exhibit worsened mental health conditions. Furthermore, HCWs who reported encountering suspected or confirmed COVID-19 cases were more likely (OR 1.55, 95% CI 1.41-1.70; $P<.001$) to exhibit worsened mental health conditions.

**Access to PPE**

HCWs who never had access to PPE such as respirators, eye protection, and face shields exhibited the highest likelihood of worsened mental health conditions than those who always had access to PPE.
access to such PPE. For instance, HCWs who never had access to respirators were more likely (OR 1.31, 95% CI 1.07-1.62; \( P < .001 \)) to exhibit worsened mental health conditions than those who did not need this PPE. Similar trends were exhibited with access to eye protection (OR 1.51, 95% CI 1.17-1.96; \( P < .001 \)) and face shields (OR 1.41, 95% CI 1.05-1.92; \( P = .02 \)).

**Discussion**

**Principal Findings**

The statistical analysis in this study found that HCWs who are younger than 35 years old were found to be more likely to exhibit worsened mental health conditions than HCWs aged 35-44 years. Male HCWs were less likely to exhibit worsened mental health conditions than female HCWs. Immigrant HCWs were also less likely to exhibit worsened mental health conditions than nonimmigrant HCWs. In contrast, HCWs working in Alberta had a higher likelihood of worsened mental health conditions than HCWs working elsewhere (Atlantic provinces, Quebec, Manitoba, Saskatchewan, Ontario, British Columbia, and Northern Territories). Frontline workers were more likely to exhibit worsened mental health conditions than nonfrontline HCWs. Part-time HCWs were less likely to exhibit worsened mental health conditions than full-time HCWs. HCWs who reported encountering COVID-19 cases were more likely to exhibit worsened mental health conditions than HCWs who reported no contact with the disease. As for PPE, HCWs who never had access to respirators, eye protection, and face shields were more likely to exhibit worsened mental health conditions than those who always had access to the same PPE, respectively.

**Comparison With Prior Work**

Previous research has shown that shock events can result in psychological trauma to HCWs [19-21]. To the best of our knowledge, our study is the first to assess the association of a comprehensive variety of factors (demographics, occupational factors, and access to PPE) independently with the mental health conditions of HCWs.

With regard to demographic factors, a study found sex to be associated with worsened mental health conditions of HCWs. Female medical staff exhibited a higher incidence of severe anxiety than their male counterparts [22]. Sex was also found as a predictor of increased anxiety and distress in a study that found female HCWs more vulnerable to such conditions than male HCWs [23]. Our study findings corroborate these previous findings. Similarly, for age, our results demonstrated that older HCWs exhibited less likelihood of worsened mental health conditions. This agrees with findings from a study in the Middle East that found older workers enjoyed better mental health conditions than younger workers [24]. Older HCWs with longer work experience seemed to have handled the pandemic better than their younger counterparts. Another study found that HCWs 40 years or older were less likely to report higher anxiety during the pandemic than younger HCWs [25]. Our results confirm such findings as well.

As for occupational factors, our findings align with a cross-sectional study conducted in Oman, which found frontline workers to be more likely to have anxiety and sleep problems [26]. This was attributed to frontline workers’ increased awareness of the mortality rate of COVID-19 and their fear of contracting the virus. Similar findings were reported by a study conducted in China as well [27]. Exposure to COVID-19 cases has been recognized as a risk factor associated with an increased likelihood of mental health issues in the literature. For instance, in France, female urologists working on the frontline were 1.41 times more likely to feel a degree of stress during their duties. Of this group, those who worked in a department where patients with COVID-19 are treated were 1.85 times more likely to report a degree of stress during work duties [28]. This was attributed to the workers’ fear of infection and the spread of the virus.

Direct involvement with COVID-19 care is also found to be highly associated with fear, depression, and anxiety as compared with those working under lower risk conditions [29]. The authors attributed this to the workers’ fear of bringing the virus to their families at home.

As for the work setting, contrary to our findings, the literature reports higher levels of fear and anxiety among part-time HCWs than among full-time HCWs [30]. Another study in France found part-time HCWs to exhibit a greater association with distress [31]. Given the small number of part-time HCWs in our study, caution should be observed in interpreting the results. Regarding factors related to access to PPE, access to PPE was found to be associated with better health and less stress [24]. Lack of access to PPE was found to be a major source of HCW stress in a study that surveyed emergency physicians [32]. In Canada, a study found that inadequate PPE supply is associated with increased symptoms of anxiety and depression among HCWs [33]. Workers were concerned about the ability to access sufficient PPE during work hours. These findings are in line with our results. Compared to studies in the literature, our study is the first comprehensive study that assesses the association of a diverse pool of factors including demographics, occupational, and access to PPE, with HCWs’ mental health conditions.

**Implications: Mental Health Programming**

Our comprehensive study sheds light on the association of various factors with HCWs’ mental health after the start of the pandemic. It helps in understanding the vulnerability of various HCW groups to mental health during such events. Certain groups were at substantially higher risk of exhibiting worsened mental health conditions after the pandemic, hence the need for a specialized support program to target this group. As a direct implication, the findings can be used to inform guidelines for mental health support for HCWs during future public health emergencies. Such mental health support may be directed more specifically to more vulnerable groups. Literature has pointed to the importance of an evidence-based approach to designing mental health support programs for HCWs [34,35]. For instance, a study in Alberta during the pandemic pointed to the importance of understanding HCWs’ occupational settings and mental health mitigation techniques [16]. Others have discussed the importance of support programs for HCWs during the pandemic based on need and background [36,37]. More importantly, many studies highlighted the importance of designing personalized mental health support programs based on various factors including demographic and occupational factors [38,39]. Some of these programs used mobile technology to offer mental health
support to HCWs [35,40]. As such, our study aims to inform stakeholders of such factors associated with HCW groups most vulnerable to future events such as the pandemic and health crises.

Limitations and Future Directions
This study has several limitations. First, the scope of the study was limited to assessing mental health in general without measuring different mental condition forms such as depression, anxiety, and fear. This is due to the limited data offered by the survey we used. Second, participants were asked, “ Compared to before the COVID-19 pandemic, how would you say your mental health is now?” That is, HCWs had to objectively compare their mental health state from the time they received the survey to sometime in the past. There may be some variation in how HCWs perceive this comparison. Third, data collected were at a single point in time, which limits the ability to compare mental health assessments to later periods in the pandemic and during subsequent pandemic waves. Fourth, no information about HCW medical history is collected in this survey. In future research, we recommend collecting data at different points in the given event to allow for a time-series analysis and comparison at different crucial points of the event. We also suggest the collection of distress information and whether the conditions persisted beyond the initial shockwave. HCWs’ years of work experience and ethnicity information can also be useful in future studies. We also believe medical history, particularly a preexisting history of mental health disorders, is an important factor associated with the mental health conditions of HCWs and should be considered in future studies.

Conclusions
This study investigated the factors associated with worsened mental health conditions of HCWs in Canada during the pandemic. Our study suggested the association of various factors with the likelihood of HCWs exhibiting worsened mental health conditions as compared to feeling neutral and better. In agreement with the literature, our findings concluded that younger (vs older), female (vs male), nonimmigrant (vs immigrant), full-time (vs part-time), and frontline (vs nonfrontline) HCWs living in Alberta (vs other provinces) exhibited a higher likelihood of worsened mental health conditions than those who felt neutral or better. Those who reported concerns about access to PPE also exhibited the same trend. Such findings can guide the future development of health care programming and inform mental health support planning for HCWs. COVID-19 is a shocking event that introduced uncertainty to the health care system.

Acknowledgments
We would like to thank Statistics Canada for providing the data used in this study.

Data Availability
The data that support the findings of this study is available through Statistics Canada.

Authors' Contributions
AA processed and cleaned the data and wrote the first draft of the manuscript with guidance from ZAB. BC and SD conducted iterative reviews and suggested revisions to the draft. All authors offered substantial contributions to the discussion and analysis presented in this manuscript. All authors agreed to publish the final draft of this paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Logistic model.

References


Gupta B, Sharma V, Kumar N, Mahajan A. Anxiety and sleep disturbances among health care workers in India cross-sectional online survey. JMJIR Public Health Surveill 2020;6(4):e24206 [FREE Full text] [doi: 10.2196/24206] [Medline: 33320910]


**Abbreviations**

HCW: health care worker  
OR: odds ratio  
PPE: personal protective equipment
A Web-Based Tool to Assess Social Inclusion and Support Care Planning in Mental Health Supported Accommodation: Development and Preliminary Test Study

Sharon Eager, MSc; Helen Killaspy, MBBS, PhD; Joanna C, MSc; Gillian Mezey, MBBS; Peter McPherson, DPsych; Megan Downey, BSc; Georgina Thompson, MSc; Brynmor Lloyd-Evans, PhD

Abstract

Background: Individuals with severe mental illness living in supported accommodation are often socially excluded. Social inclusion is an important aspect of recovery-based practice and quality of life. The Social Inclusion Questionnaire User Experience (SInQUE) is a measure of social inclusion that has been validated for use with people with mental health problems. Previous research has suggested that the SInQUE could also help support care planning focused on enabling social inclusion in routine mental health practice.

Objective: This study aims to develop a web-based version of the SInQUE for use in mental health supported accommodation services, examine its acceptability and perceived usefulness as a tool to support care planning with service users, determine the extent of uptake of the tool in supported accommodation settings, and develop a program theory and logic model for the online SInQUE.

Methods: This study involved a laboratory-testing stage to assess the acceptability of the SInQUE tool through “think-aloud” testing with 6 supported accommodation staff members and a field-testing stage to assess the acceptability, utility, and use of the SInQUE tool over a 5-month period. An implementation strategy was used in 1 London borough to encourage the use of the SInQUE. Qualitative interviews with 12 service users and 12 staff members who used the tool were conducted and analyzed using thematic analysis. The use of the SInQUE was compared with that in 2 other local authority areas, 1 urban and 1 rural, where the tool was made available for use but no implementation strategy was used.

Results: Overall, 17 staff members used the SInQUE with 28 different service users during the implementation period (approximately 10% of all service users living in supported accommodation in the study area). The staff and service users interviewed felt that the SInQUE was collaborative, comprehensive, user-friendly, and relevant. Although some staff were concerned that particular questions might be too personal, service users did not echo this view. Participants generally felt that the SInQUE could help identify individuals’ priorities regarding different aspects of social inclusion by prompting in-depth conversations and tailoring specific support to address service users’ inclusion goals. Some interviewees also suggested that the tool could highlight areas of unmet or unmeetable needs across the borough that could feed into service planning. The SInQUE was not used in the comparison areas that had no implementation strategy.

Conclusions: The online SInQUE is an acceptable and potentially useful tool that can be recommended to assess and support care planning to enable social inclusion of people living in mental health supported accommodation services. Despite this, uptake rates were modest during the study period. A concerted implementation strategy is key to embedding its use in usual care, including proactive endorsement by senior leaders and service managers.
social inclusion; supported accommodation; mental health; digital health; care planning

Introduction

Background

Social inclusion refers to an individual’s ability to participate in important societal activities and their sense of community belonging [1,2]. Someone may feel socially excluded if they do not have opportunities for societal involvement and integration, often because of external factors that are beyond their control [3]. Social exclusion is a multifaceted continuum [2], typically signified by poverty, unemployment, inequality, and poor health [4].

People with serious mental illness are thought to be among the most socially excluded groups in society [5]. Individuals with this type of mental health problems often have smaller and less satisfying social networks [6], lower household income [7], and lower levels of employment [8,9] and experience more criminal and violent victimization [10,11] than those in the general population. Social exclusion can be conceptualized as both a cause and a consequence of mental illness [12]. Furthermore, greater social inclusion is associated with better quality of life and lower levels of loneliness among those with severe mental illness, suggesting that social exclusion is an important area for mental health practitioners to try to address [13,14].

Mental health supported accommodation services provide care and support to individuals with particularly severe and complex mental health problems as a way of supporting recovery in the community [15]. It is estimated that there are approximately 100,000 people living in mental health supported accommodation in England. Services are typically staffed by support workers, with additional specialist clinical input provided by National Health Service (NHS) community mental health teams [16]. In England, three main types of supported accommodation are provided: (1) residential care homes for those with the highest needs that comprise 24-hour–staffed communal facilities where placements are not time limited, with meals, supervision of medication, cleaning, and activities provided to service users; (2) supported housing services that provide shared or individual, self-contained, and time-limited tenancies with staff on-site up to 24 hours a day to assist service users in gaining skills to move on to less supported accommodation; and (3) floating outreach services that provide visiting support for a few hours per week to people living in permanent, self-contained, and individual tenancies with the aim of reducing support over time to zero [16].

Service users living in supported accommodation are often socially isolated, with low levels of employment and little involvement in civil and political processes [17]. Many report feeling lonely and isolated and experiencing a high level of stigma that causes them to become more socially isolated [18]. There is evidence that users of mental health supported accommodation services report a variety of unmet needs, such as accessing employment opportunities and forming intimate relationships [19,20]. However, relatively little research has been conducted to determine the precise needs of service users living in supported accommodation [21], and a greater focus on this group is needed to identify and implement interventions that are likely to be the most useful for them [22].

Supporting service users to work toward desired goals and community engagement is highly congruent with recovery-based practice in mental health. Recovery-based practice recognizes and builds on service users’ strengths and promotes empowerment through collaboration between them and staff to identify and work toward specific goals [23]. Many of the identified goals are markers of social inclusion, such as employment, social network development, and participation in community activities [24]. There is qualitative evidence from a large national research program suggesting that staff working in mental health supported accommodation services operate with a considerable degree of recovery orientation [24,25], and the more recovery-orientated these services are, the more likely people are to move on successfully to more independent settings [26].

People living in mental health supported accommodation have expressed a strong preference for individually tailored services that offer choice and promote autonomy, consistent with a recovery-based approach [18]. Patient-reported outcome measures have been recommended to inform this individualized approach by directly capturing service users’ perspectives on constructs such as goal attainment, quality of life, and social inclusion [27]. Such measures enable service users to make informed decisions about their own support and care planning, in line with World Health Organization recommendations for recovery-based practice in community care provision [28]. Resources delivered across web-based platforms, particularly those that offer guided support, have been established as accessible, acceptable, and effective for use by participants with severe mental illness [29,30]. A tailor-made web-based assessment tool, the Quality Indicator for Rehabilitative Care for supported accommodation, has also been successfully used by managers of supported accommodation services, suggesting that these settings have the required resources and expertise to implement online measures [31].

The Social Inclusion Questionnaire User Experience (SInQUE) was developed as a measure of social inclusion for individuals with severe mental illness [32]. The measure has been validated across a range of mental health populations, has established reliability, is considered acceptable to service users, and has been proposed as being potentially cross-culturally suitable [32-34]. To date, the SInQUE has been used solely in offline research contexts. However, stakeholder feedback from a previous study testing the SInQUE indicated that the measure may be useful in clinical practice to assess social inclusion, facilitate important conversations with service users, and guide care and support planning [34]. Furthermore, a consistent research recommendation from the developers of the SInQUE
tool has been to investigate whether the measure has utility as a care-planning tool to promote social inclusion in routine mental health practice [13,34].

This study aimed to develop a web-based version of the SInQUE for use in mental health supported accommodation services. We sought to examine the acceptability and perceived usefulness of this tool among supported accommodation staff and service users as means to assess their needs for greater social inclusion and promote care planning.

Aims
The study aims were as follows:

1. To develop and refine a web-based version of the SInQUE social inclusion assessment tool tailored for use in mental health supported accommodation settings.
2. To investigate the acceptability and perceived utility of the online SInQUE tool among supported accommodation staff and service users.
3. To determine the extent of uptake of the tool in supported accommodation settings with and without a locally developed implementation strategy to support its use.
4. Informed by the study findings, to develop a program theory and logic model for the online SInQUE specifying its anticipated outcomes, the mechanisms through which they may be achieved, and contextual factors affecting the use and experience of the SInQUE.

Methods

Study Design
This study comprised two stages conducted in 1 inner London borough:

1. A laboratory-testing stage to assess initial acceptability of the tool and develop it through “think-aloud” testing and semistructured interviews with supported accommodation staff.
2. A field-testing stage to assess wider acceptability, feasibility, and use of the tool. Semistructured interviews were conducted with staff and service users who had used the online SInQUE during this stage.

The 5-month field-testing stage was supported by a local implementation strategy developed in collaboration with local service leads to support the use of the online SInQUE by supported accommodation staff in the participating London borough. The online SInQUE was also made available to supported accommodation services by local service leads in 2 other areas without any accompanying implementation strategy.

Description of the SInQUE Tool
The web-based version of the SInQUE [35] can be used to assess social inclusion and inform support and care planning for people with mental health problems. It is designed to be used by staff as part of routine care planning to be completed collaboratively with service users. It can be used on a computer, tablet, or mobile device. Staff are required to register for an account on the SInQUE site using their work email address and details of their organization and can then use the tool for free. No personal data identifying service users are logged or stored on the SInQUE platform. The online SInQUE generates a unique reference number for each new service user, which is retained by the staff member completing the assessment for future reference and to link any repeat assessments.

The online tool developed for laboratory testing in our study included the 46-item version of the validated SInQUE social inclusion questionnaire, which was refined following stakeholder feedback at the end of the previous measure development study [34]. For this study, we removed 1 question from the SInQUE that asked whether the respondent was living alone as this was considered redundant for people living in residential care and supported housing. The SInQUE’s psychometric properties have been established among people with a range of mental health problems receiving input from community mental health services in previous studies [13,32,34]. Although the removal of a question from the original SInQUE questionnaire compromised its established psychometric properties, this minor adaptation is unlikely to have disrupted them substantially. We wanted to minimize changes to maintain the online SInQUE’s similarity to the validated measure and did not aim to make further significant refinements to the content of the tool. Instead, we wanted to gain feedback on its acceptability and feasibility for use in its digital format among staff and users of mental health supported accommodation services to assist in care and support planning.

The online SInQUE questionnaire yields a total score of 0 to 75, with a higher score indicating greater social inclusion. The questions and subscale scores are grouped into 9 different areas of social inclusion: leisure, social relationships, religious and cultural activities, education and employment, transport, health, crime victimization, home life and housing, and civic duties. These areas cover the 5 social inclusion domains of the validated SInQUE (social integration, productivity, consumption, access to services, and political engagement), but the aforementioned 9 areas were considered more immediately understandable for use in practice.

Using the service user’s responses, the online SInQUE generates a list of areas in which the person has said that they would like to be more socially included. It then offers a prompt for the service user and staff member to collaboratively select up to 3 priority areas that they would like to integrate into the person’s support plan. Once the assessment is completed, a summary report is generated. If the assessment is repeated with the same service user in the future, this report will also display changes in their social inclusion over time. The tool can generate management-level summary reports for each organization that is registered with it and commissioner-level summary reports of services using the tool across an entire area (such as a London borough). Multimedia Appendix 1 provides a full description of the SInQUE using the TIDieR (Template for Intervention Description and Replication) checklist [36].

Setting
This study took place in mental health supported accommodation services across 1 inner London borough. There are 21 such services in the borough run by 6 different voluntary sector organizations. They offer varying degrees of support to >270
service users who are also supported by local NHS secondary mental health services. In the borough, there are approximately 24 service users living in residential care, 159 living in supported housing, and 89 who receive floating outreach support. Supported housing services offer 24-hour support to 119 people and “9 to 5” support to 40 individuals.

**Laboratory-Testing Stage: Recruitment, Data Collection, and Analysis**

In total, 6 supported accommodation staff members were recruited to provide important insights of the online SInQUE tool. We discussed the study with service managers working in 3 different services and asked them to nominate 2 staff members each from their service who were interested in taking part. Participants were purposively sampled to include staff working in floating outreach support, 24-hour supported housing, and residential care. We asked the managers of each of the 3 services to ask for volunteers from their staff teams. Interviews were arranged with the first 2 staff members identified by the manager of each service.

Data were collected between January 2022 and February 2022. All 6 think-aloud interviews with staff members were conducted and recorded using Microsoft Teams (Microsoft Corp). The researcher first discussed the information sheet with each participant and gave them the opportunity to ask questions. Following this, participants’ consent was verbally collected and audio recorded separately from the main part of the interview. Participants were then asked to fill out a short online form providing their demographic information before beginning the interview.

We conducted “think-aloud” testing of the online tool with staff using a semistructured topic guide developed by the study team ([Multimedia Appendix 2](#)). Following a process used previously in developing web-based tools [37], participants were asked to complete set tasks using the online SInQUE tool while providing a continuous commentary on their thoughts. They were asked to open the SInQUE website, register for an account, and complete an assessment as they would with a service user. At all stages, they were prompted to share their thoughts as they navigated the website and offer their initial impressions on how easy it was to understand and use and its potential suitability for their work. Once participants had completed the questionnaire, the researcher asked broader questions about their experience using the tool and any ways in which it could be improved. Throughout these interviews, participants were asked to focus on their experience using the SInQUE tool rather than offering specific feedback on individual SInQUE items. This was because we did not intend to make substantial modifications to the SInQUE questions to maintain their scope and similarity to those of the validated SInQUE measure.

Identified problems and suggestions for improvements to the online tool were collated by the researcher following the interviews. They were then reviewed by the study team, decisions about refinements to the online SInQUE were agreed upon, and the tool was revised accordingly.

**Field-Testing Stage: Recruitment, Data Collection, and Analysis**

**Local Implementation Strategy**

The revised version of the tool was made available for use in mental health supported accommodation services across the participating inner London borough. We iteratively developed and implemented a strategy to encourage and support its use by supported accommodation staff in the borough over a 5-month period beginning on May 11, 2022. This implementation strategy was informed by consultation with supported accommodation service managers and clinicians working in the Islington community mental health rehabilitation team and by individual interviews conducted with supported accommodation service users and staff.

**Interviews With Field-Testing Participants**

**Participants and Recruitment**

Individual interviews with service users (n=12) and staff (n=12) who tried out the SInQUE tool were conducted from late May 2022 to September 2022. This number was chosen to explore the views of staff and service users from a variety of supported accommodation types and service providers. Following our implementation strategy, we asked staff members to alert the study researcher once they had tried the tool in practice. Any staff member or service user in supported accommodation who tried the SInQUE tool was eligible to participate in an individual interview.

Once a staff member informed the study researcher that they had tried the tool, we asked them whether they would like to participate in an individual interview about their experience. We also invited staff to pass on information about the study to the service users with whom they had used the tool and ask them whether they would like to participate in an interview about their experience. If the service user was interested in taking part, the researcher communicated with them either directly or through the staff member they had completed the SInQUE assessment with depending on their preference. Toward the end of the recruitment stage, we recruited the final few staff members and service users purposively to ensure that participants were from a range of supported accommodation types and provider organizations.

One service user interview and 1 staff member interview were conducted online via Microsoft Teams; all other interviews were carried out in person according to participants’ preferences. In-person interviews were conducted by the study researcher at the staffed supported accommodation sites, aside from 1 interview with a service user receiving floating outreach support, which was conducted at their home.

**Measures and Procedures**

The researcher first discussed the information sheet with the participants and gave them the opportunity to ask questions about the study. For in-person interviews, informed consent was collected via a paper consent form; for online interviews, verbal consent was audio recorded. Participants were then asked to answer brief demographic questions about themselves and their associated services. Following this, the researcher asked
each participant questions about their experience using the SInQUE; whether there were any ways in which it could be improved; the appropriateness of the online tool for use in their work; and what impacts, if any, they thought it might have on care provision and service users’ experience. The interview topic guides (one for staff participants and one for service user participants) were developed by the study team as semi-structured interviews—they are provided in Multimedia Appendix 3. In-person interviews were recorded using a digital voice recorder; online interviews were recorded on Microsoft Teams. Interview audio recordings were transcribed by a professional transcription company with which University College London (UCL) had a data-sharing and privacy agreement. Interview transcripts were then checked by the study researcher for accuracy. Any potentially identifiable text was anonymized. The resulting cleaned transcripts were then securely stored on the UCL university system.

Analysis

The analysis of the interviews comprised 2 stages. First, the study researcher noted any problems experienced by participants and recorded improvements to the online SInQUE they suggested. These issues and the suggested changes were reviewed by the study team, as in the previous laboratory-testing phase. Minor modifications to the online SInQUE were agreed upon, and we made adjustments to the tool in line with this.

Second, transcripts were uploaded to NVivo (version 12; QSR International) for qualitative analysis. As we aimed to develop a program theory for the online SInQUE intervention, we initially coded data into a deductively derived framework that used an intervention-context-actor-mechanism-outcome (ICAMO) configuration, with each component of this ICAMO framework representing a primary theme [38]. Within each of these 5 primary themes, we inductively derived subthemes from the data using thematic analysis. The initial coding was conducted by the lead author (SE) and was then reviewed and adjusted collaboratively by the study team. This included gaining lived-experience perspectives from a researcher with experience of mental health service use (JC) and clinical insights from a senior clinical academic working in the participating borough as a consultant rehabilitation psychiatrist supporting service users who live in supported accommodation (HK). The team brought in further perspectives from those with backgrounds in social work (BLE), clinical psychology (PM), and forensic psychiatry (GM) and from the community rehabilitation team in the borough (MD).

Data Use Monitoring

Data on the uptake and use of the online SInQUE tool were collected from the online SInQUE informatics for the 5-month field-testing period from May 11, 2022, to October 11, 2022. At the start of this period, the study team also contacted local mental health service rehabilitation and housing leads in 2 other areas: another inner London borough and a rural county in the west of England. These service leads contacted local supported accommodation managers and invited them to use the online SInQUE in their service if they wished. The tool was made available to 7 supported accommodation services in the London borough and 10 in the rural county. No further encouragement to use the tool or implementation support was provided. This allowed us to monitor uptake and use of the tool in 2 areas without an associated implementation plan, thus making inferences about the necessity and impact of the strategy we developed.

Logic Model Development

The study team developed a preliminary logic model for the online SInQUE in planning this study. We used the findings of the aforementioned research activities to review and refine this logic model and develop an updated theory about the potential outcomes for service users and organizations from using the online SInQUE; mechanisms through which these outcomes are achieved; and factors influencing the uptake, experience, and impact of the online tool. Factors were related to (1) the intervention itself, (2) the characteristics and attitudes of staff and service users using the online SInQUE, and (3) the broader organizational and societal context. This was summarized in a logic model in the form of an “ICAMO map” [38], which was developed and refined iteratively through discussion with the study team.

Ethical Considerations

The initial laboratory-testing phase of this study (Supporting social inclusion for people with serious mental illness living in supported housing [SUSHI] phase 1) was approved by the UCL Research Ethics Committee (REC) on June 18, 2021 (REC reference 6711/002). The subsequent field-testing phase (SUSHI phase 2) was approved by the London – Camden and Kings Cross NHS REC on November 4, 2021 (REC reference 21/LO/0657). Written or audio-recorded informed consent was obtained from all participants before they took part, and they were clearly informed that they could opt out of the study at any time. All the study data were carefully deidentified. Service user participants were offered a £20 (US $25.14) shopping voucher to thank them for their time.

Results

Participants

We recruited 6 supported accommodation staff members for the “think-aloud” interviews during the laboratory-testing stage. We recruited a further 12 staff members and 12 supported accommodation service users for the individual interviews as part of the field-testing stage. Participant characteristics for both stages are summarized in Table 1.
<table>
<thead>
<tr>
<th>Participant characteristics</th>
<th>Laboratory testing</th>
<th>Field testing</th>
<th>Service users (n=12), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2 (33)</td>
<td>5 (42)</td>
<td>11 (92)</td>
</tr>
<tr>
<td>Female</td>
<td>4 (67)</td>
<td>6 (50)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Age group (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-30</td>
<td>0 (0)</td>
<td>8 (67)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>31-50</td>
<td>2 (33)</td>
<td>2 (17)</td>
<td>3 (25)</td>
</tr>
<tr>
<td>≥51</td>
<td>4 (67)</td>
<td>1 (8)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian/Asian British</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Black/Black British</td>
<td>3 (50)</td>
<td>2 (17)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>White/White British</td>
<td>2 (33)</td>
<td>8 (67)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Mixed/multiple ethnic groups</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Other ethnic background</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Sexual orientation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>N/A</td>
<td>N/A</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Gay/lesbian</td>
<td>N/A</td>
<td>N/A</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>N/A</td>
<td>N/A</td>
<td>2 (17)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>N/A</td>
<td>N/A</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Type of supported accommodation lived or worked in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floating outreach support</td>
<td>2 (33)</td>
<td>2 (17)</td>
<td>2 (17)</td>
</tr>
<tr>
<td>9-to-5 supported housing</td>
<td>2 (33)</td>
<td>1 (8)</td>
<td>2 (17)</td>
</tr>
<tr>
<td>24-h supported housing</td>
<td>2 (33)</td>
<td>7 (58)</td>
<td>8 (67)</td>
</tr>
<tr>
<td>Residential care</td>
<td>0 (0)</td>
<td>2 (17)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Length of time worked or lived in supported accommodation (y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>1 (17)</td>
<td>7 (58)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>2-5</td>
<td>3 (50)</td>
<td>3 (25)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>6-10</td>
<td>1 (17)</td>
<td>2 (17)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>≥10</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
</tbody>
</table>

aN/A: not applicable; staff were not asked about their sexual orientation.

**Changes Made to the SInQUE**

Following phase 1 laboratory testing and phase 2 field-testing, suggestions that participants made for how the tool could be improved were collated and reviewed by the team. Accordingly, adjustments were made to the online SInQUE after each stage, an overview of which can be found in Table 2. This addressed aim 1 of this study.
Table 2. Changes made to the online Social Inclusion Questionnaire User Experience (SInQUE) following phase 1 and phase 2 testing. All changes were made following the initial laboratory-testing stage unless indicated otherwise.

<table>
<thead>
<tr>
<th>Section of the SInQUE affected</th>
<th>Explanation of the problem</th>
<th>Resolution</th>
<th>Justification for the change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Registration and use changes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The online SInQUE home page</td>
<td>Some staff members and service users suggested developing additional materials to help explain the SInQUE(^a).</td>
<td>Developed a guidance manual for service managers and commissioners with information on using the SInQUE as well as an informational leaflet and poster aimed at service users about the SInQUE.</td>
<td>It is easier for managers, commissioners, staff, and service users to understand and use the SInQUE.</td>
</tr>
<tr>
<td>The initial page where staff members are asked to register for the SInQUE</td>
<td>Staff were asked to enter the organization they worked for in a free-text box. Some found it confusing to know which organization name they should enter.</td>
<td>Changed response options to a fixed-response drop-down menu with all housing providers in the borough and an “other” free-text option.</td>
<td>Allows for compilation of service-level data and is easier for staff to navigate.</td>
</tr>
<tr>
<td>The page where staff members enter details to set up a new SInQUE assessment</td>
<td>Some staff members thought that further information on the exact purpose of the questionnaire and how it should be administered would be useful in the introductory paragraph describing the assessment.</td>
<td>Additional guidance on how the questionnaire should be administered was added to the introduction paragraph of the SInQUE.</td>
<td>Important contextual information for the questionnaire was explicitly clarified.</td>
</tr>
<tr>
<td>The page where staff members enter details to set up a new SInQUE assessment</td>
<td>Some staff members found the wording of the following question—“Please select the type of accommodation in which the service user is living from the list below”—to be ambiguous and confusing.</td>
<td>Changed the wording of the question to the following: “Please select the type of housing support the service user receives from the list below.”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td><strong>Changes to the wording of SInQUE questions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The section covering “leisure” questions</td>
<td>Some staff members felt that it was unclear what the “Other” option meant in the context of question 3f: “Over the past year have you been to...Other?”</td>
<td>Changed the wording of this subquestion to the following: “Other leisure activity?”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td>The section covering “leisure” questions</td>
<td>Some staff members and service users felt that question 7—“Do you spend time in pubs or cafés?”—was worded in a way that was potentially inappropriate for people who do not drink alcohol(^a).</td>
<td>Changed the wording of the question to the following: “Do you go out for a coffee/drink (e.g. to a café or pub, etc) at least once a week?”</td>
<td>The clarity and appropriateness of the question improved.</td>
</tr>
<tr>
<td>The section covering “social” questions</td>
<td>Some staff members were unsure whether question 9—“How many people, outside those in your care team, could you confide in?”—related to a professional or personal care team.</td>
<td>Changed the wording of the question to the following: “How many people, outside the workers in your care team, could you confide in?”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td>The section covering “home life/housing” questions</td>
<td>Some staff members thought that question 36—“What kind of accommodation do you live in?”—was worded ambiguously.</td>
<td>Changed the wording of the question to the following: “What type of housing support do you receive?”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td>The section covering “home life/housing” questions</td>
<td>A statement alerting users that question 38 had been omitted from the online SInQUE, which read the following—“Not relevant for supported accommodation contexts - omitted.”—was confusing.</td>
<td>Changed statement to the following: “Question omitted, not included in the online SInQUE.”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td><strong>Changes to the SInQUE summary outputs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The SInQUE summary report</td>
<td>Some staff members found the spider graph to be confusing to interpret as the numbers summarizing scores in each section were not standardized and, therefore, it was difficult to tell which domains scored lower than others.</td>
<td>Simplified the spider graph to show percentage of total score in the graph instead of frequency.</td>
<td>It was easier to interpret the graph as sections with different totals became standardized.</td>
</tr>
<tr>
<td>The SInQUE summary report for multiple assessments completed with the same service user</td>
<td>In the section summarizing scores across multiple time points, staff members thought that a visual depiction of this comparison would be useful.</td>
<td>Made comparative bar charts of multiple scores across time with the same service user available on the summary report.</td>
<td>It was easier to understand and relay the results.</td>
</tr>
</tbody>
</table>

\(^a\)Changes were made following the field-testing stage.
Overall, the changes made to the online SInQUE were relatively few and minor. Following initial laboratory testing, additional information and guidance for users was added, and minor revisions to the wording of questionnaire items were made to improve clarity. Modifications to the visual representation of scores in the summary reports were also made to aid ease of interpretation. During the field-testing stage, very few suggested changes to improve the usability of the online SInQUE were made by staff or service user participants. Further changes made at this stage included a minor wording adjustment to one question to ensure its cross-cultural appropriateness. Changes were made exclusively to the web-based version of the SInQUE and did not affect the existing SInQUE measure. We also developed an additional guidance document for managers and commissioners and an informational leaflet and poster about the SInQUE.

A few participants suggested substantial modifications to the structure and wording of individual items in the SInQUE that were not implemented by the study team. These decisions were made to preserve the broad scope and logical flow of the tool. We also declined to action some suggestions that were outside the remit of the SInQUE tool, such as adding more or free-text response options to some questions. However, where these suggestions indicated important potential barriers to using the SInQUE, they were noted and integrated into the qualitative analysis and logic model development. A summary of all comments and suggestions that were proposed but not implemented after review by the team can be found in Multimedia Appendix 4.

Interview Thematic Analysis

Overview

The interviews were analyzed using thematic analysis. Primary themes were deductively imposed according to each core element of the ICAMO model: intervention, context, actors, mechanisms, and outcomes [38]. Subthemes were inductively analyzed within each of these primary themes. The resultant thematic framework considering the perceived utility and acceptability of the online SInQUE and addressing aim 2 of the study is presented in Textbox 1. The themes are summarized in the following sections with a selection of illustrative quotes. Multimedia Appendix 5 provides further illustrative quotes for each subtheme.
**Textbox 1.** Summary of the thematic framework (intervention-context-actor-mechanism-outcome themes and inductive subthemes).

### Intervention: combination of program elements or strategies designed to produce behavior changes or improve health status among individuals

- The online Social Inclusion Questionnaire User Experience (SInQUE):
  - Promotion of positive, collaborative discussion
  - Comprehensive and novel questions
  - Ability to repeat over time
  - User-friendly design:
    - Easy-to-navigate website
    - Quick to complete
    - Fixed-response questions
    - Offers options to choose from
    - Web-based format

### Context: salient conditions that are likely to enable or constrain the activation of program mechanisms

- Relevance of the SInQUE to staff role
- Inconsistency in current assessments used across services
- Absence of comparably specific assessments
- Emergence from the pandemic

### Actors: the individuals, groups, and institutions who play a role in the implementation and outcomes of an intervention

- Staff:
  - Professional knowledge and skills
  - Professional boundaries
- Staff (service user views about staff):
  - Trusting relationship
  - Proactivity in offering guidance and support
- Service users:
  - Familiarity and comfort with the questions
  - Individual language and cultural differences of service users
- Service users (staff views about service users):
  - Engagement in the assessment
  - Existing mental health needs

### Mechanisms: any underlying determinants or social behaviors generated in certain contexts

- Using the online SInQUE can accomplish the following:
  - Boost service user proactivity and confidence
  - Identify service users’ priorities on social inclusion
  - Prompt novel, personal conversations
  - Monitor changes in social inclusion over time
  - Identify gaps in support available within the organization and local community

### Outcomes: behavior changes that follow the immediate knowledge change (intermediate) and changes such as patients’ health status and impact on community and health system (long term)

- Intermediate:


**Intervention**

In general, staff and service user participants felt that the tool was user-friendly and collaborative. Many noted the ability to repeat the assessment and the web-based format as being particularly useful and felt that the website was easy to navigate. The short length of the assessment was also discussed as an important advantage, with both staff and service users commenting that it felt quick to fill out. Participants noted that, despite the short assessment length, it still offered a range of interesting, positive, and sometimes unfamiliar questions that felt comprehensive and useful to discuss:

*I think that it wasn’t too just baseline, it was a little bit more than that and I think that’s good. Because it gives the option of, “Okay, you don’t want something, how can we improve and what is it that you do want that could help you while you’re in our service?”* [212; staff member]

Interviewees felt that the user-friendliness was aided by accessible questions that were straightforward for service users to answer and that were cross-culturally appropriate for individuals from different ethnic backgrounds. Although some participants from both groups felt that the fixed-response options for the questions were limited, certain staff members thought that this made the questionnaire more accessible to service users who may otherwise struggle with engagement.

**Context**

Staff members largely felt that the tool was fitting and relevant to their role in helping support service users, and most did not already use assessments that were highly similar to the SHnQUE. Certain staff members highlighted a lack of continuity of support workers within their service and noted that this often made it difficult to build rapport with service users. Some also commented on an inconsistency in assessments used across different services (in the local context, where 6 different provider organizations provided supported accommodation services across the borough). They noted that individual providers currently make their own recommendations on the tools that staff should use:

*If it was a standard central assessment that we do in all supported housing, that’s similar, like this for example, it might be beneficial in the long run. But each company has their own policy around it.* [205; staff member]

One staff member noted that the tool felt particularly relevant following the COVID-19 pandemic as a means to promote engagement among service users after a period of likely sustained social isolation.

**Actors**

There were 2 key actors to consider in the application of the assessment: the staff members who asked the questions and the service users who responded to them.

One staff member felt that the assessment was not particularly relevant in the context of their work in a residential care service, where they had an established relationship with service users and already knew much of the queried information about them. However, this was an outlying view. Although most staff members thought that the online tool could be suitable and useful for their work, they emphasized the importance of using their professional knowledge and skills to pick when and for whom the assessment would be appropriate. They suggested that service users acutely struggling with their mental health may find it difficult to maintain concentration and engagement with the questionnaire and others may feel that the assessment is not relevant to them.

Staff members also raised the importance of maintaining professional boundaries with service users, and some expressed a concern that certain questions may feel invasive or uncomfortable for service users to answer:

*I think there was one quite private like about if they’re in a relationship or something, and that was the only question that made me feel a bit like I’m asking something very personal about a relationship. Because they might not want to say that.* [210; staff member]

However, service user participants did not express any similar concerns about intrusive questions. They generally indicated that they felt comfortable with the assessment and that they were used to answering personal questions. Both staff and service users highlighted trust between those performing the assessment as a key factor in promoting engagement with such questions.

Both respondent groups highlighted the cultural diversity of service users within supported accommodation, and many noted that the tool felt appropriate for those from a variety of religious and ethnic backgrounds. Some service users commented on the importance of the staff members being proactive and taking the time to go through the assessment in detail with them, particularly individuals for whom English was a second language, as further explanation was required for some questions. Various staff members suggested that being provided with information about the assessment and its purpose specifically aimed at service users, for instance, a guidance
leaflet, would be helpful for them to convey the essential information about the assessment.

**Mechanisms**

Both participant groups discussed how the tool may boost confidence and proactivity for a wide range of service users by highlighting specific, achievable ways in which an individual can improve their social inclusion. They also noted how the assessment encourages service users to open up and enables more profound conversations between them and staff members:

*Yes I found it really interesting, so like because it’s not really topics I would actually talk about. So it gave me a bit of enthusiasm to talk about some of the questions.* [409; service user]

Both groups suggested that it might be particularly useful during key working sessions as a means to get to know an individual better and identify their support preferences soon after moving into supported accommodation. Participants also noted the value of repeating assessments over time, suggesting that this could be a potentially encouraging way to demonstrate service user progress and identify gaps in available support. The most frequently suggested time between assessments was 1 to 3 months, with up to 6 months mentioned as a potential maximum gap.

**Outcomes**

Interviewees discussed the short- and long-term outcomes that they felt the tool could offer. They discussed how the tool enabled targeted and relevant support that prioritized the service user’s interests. Both groups also mentioned the potential for the tool to improve the relationship and understanding between service users and staff members:

*It asks questions where maybe like for your support worker to get a better understanding of you, like even though the immediate thing is highlight areas you can work on, it gives a general overview of how you are.* [410; service user]

Some staff members also discussed how prolonged use of the tool could highlight the additional borough-level support that may be needed to improve certain gaps in support and could also promote service user recovery toward the goal of more independent accommodation.

**Implementation Strategy**

Our implementation strategy was developed to encourage the use of the SInQUE in the supported accommodation services in the borough and was updated through consultation with clinical staff working in the borough’s community rehabilitation team and supported accommodation service managers. Our strategy was further informed by feedback from staff and service user participants during both stages of the study.

Each part of the strategy was developed to target an identified potential barrier to staff using the online SInQUE with service users. Subsequently, we mapped each component of the strategy to the 3 broad domains of the Capability, Opportunity, and Motivation–Behavior framework of behavior change [39] to describe whether each element of the strategy was intended to increase the staff’s capability, opportunity, or motivation to use the online SInQUE. The complete implementation strategy and the Capability, Opportunity, and Motivation–Behavior domain that each component addressed are outlined in Table 3. Strategies were related to enlisting leadership support to encourage staff to use the SInQUE, providing technical guidance and assistance with using the online tool, and developing bespoke summary output reports to reinforce use and increase the organizational benefits of using the SInQUE.
Table 3. Summary of the implementation strategy to support the use of the online Social Inclusion Questionnaire User Experience (SInQUE).

<table>
<thead>
<tr>
<th>Activity</th>
<th>Implementation goal being addressed</th>
<th>COM-B(^{a}) domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study research assistant (SE) and clinical research staff member (MD) visit all supported accommodation services to introduce the SInQUE tool to staff and offer guidance on its use.</td>
<td>Increase awareness of the SInQUE tool among supported accommodation staff and respond to any of their concerns or other problems.</td>
<td>• Motivation • Opportunity • Capability</td>
</tr>
<tr>
<td>NHS(^{b}) community rehabilitation team manager and clinical researcher (MD) review supported accommodation caseloads to identify suitable service users for a SInQUE assessment and ask their key worker to complete an assessment with those service users.</td>
<td>Lack of accountability after initially asking all staff members to use the SInQUE and staff hesitancy over which service users would be suitable for an assessment</td>
<td>• Motivation</td>
</tr>
<tr>
<td>NHS community rehabilitation team manager contacts supported accommodation managers to ask them to support and encourage the use of the SInQUE by identified key workers within a given time frame.</td>
<td>Lack of supported accommodation management prioritization for staff to use the SInQUE</td>
<td>• Motivation • Opportunity</td>
</tr>
<tr>
<td>Local authority commissioners contact all supported accommodation managers to encourage the use of the SInQUE within services in the borough.</td>
<td>Lack of supported accommodation management prioritization for staff to use the SInQUE</td>
<td>• Motivation</td>
</tr>
<tr>
<td>Study lead (BLE) and research assistant (SE) attend local Housing Forum meetings to update on the study and encourage the use of the SInQUE among all managers and staff members present.</td>
<td>Increase visibility and awareness of the SInQUE among supported accommodation managers and respond to any of their concerns or other problems.</td>
<td>• Motivation • Opportunity</td>
</tr>
<tr>
<td>Study research assistant (SE) offers supported accommodation staff technological support with SInQUE registration and use.</td>
<td>Uncertainty about how to manage the technical process of using the SInQUE</td>
<td>• Capability</td>
</tr>
<tr>
<td>Study team develops and circulates a leaflet about the SInQUE for supported accommodation staff to give to service users to help explain the purpose of an assessment.</td>
<td>Uncertainty among some supported accommodation staff members about how best to explain the purpose of the SInQUE and engage service users</td>
<td>• Capability</td>
</tr>
<tr>
<td>Study team sends summary reports to service managers outlining use of the SInQUE and highlighting the areas of social inclusion that are most frequently prioritized and addressed in their service(^{c}).</td>
<td>Increased awareness among supported accommodation managers of the value offered by the SInQUE for service planning to encourage them to prompt staff to use it.</td>
<td>• Motivation</td>
</tr>
<tr>
<td>Study team sends summary reports to local authority commissioners outlining which services have used the SInQUE the most and highlighting the areas of social inclusion that are most frequently prioritized and addressed across all services in the borough(^{c}).</td>
<td>Increased awareness of commissioners of the value offered by the SInQUE for service planning and commissioning to encourage them to prompt services to use it.</td>
<td>• Motivation</td>
</tr>
</tbody>
</table>

\(^{a}\)COM-B: Capability, Opportunity, and Motivation–Behavior.

\(^{b}\)NHS: National Health Service.

\(^{c}\)These actions were planned with service managers and commissioners but not carried out during the 5-month implementation period because of the small number of completed SInQUE assessments.

Usage Data

In total, 27 staff members in the inner London borough registered for an account with the online SInQUE. Of the 27 staff members who registered, 17 (63%) from 6 different supported accommodation providers started or completed a SInQUE assessment with at least 1 service user. This resulted in 30 completed SInQUE assessments with 28 service users in the borough. This represented just >10% of the estimated total number of service users living in supported accommodation in the borough. Of the 28 service users, 4 (14%) were from residential care, 19 (68%) were from 24-hour supported housing, 3 (11%) were from 9-to-5 supported housing, 1 (4%) was from floating outreach services, and 1 (4%) was registered as “other” accommodation type. One staff member from 1 of the local authority areas where there was no specific implementation strategy registered for an account with the SInQUE; however, they did not start or complete a SInQUE assessment. This addressed aim 3 of the study.

Intervention Logic Model

On the basis of the collective study findings, we developed a logic model to summarize the processes involved in using the SInQUE and address study aim 4 (Figure 1).
The logic model was informed by the structure of an “ICAMO map” [38]. The model outlines the key aspects of the intervention (I), including its user-friendliness, its comprehensive nature, and the fact that it was based on a validated measure of social inclusion. It also indicates the potential outcomes (O) from using the SInQUE at both the individual and system levels, including improved support planning, better relationships, and provision of additional support for staff and services, which in turn may improve social inclusion and mental health outcomes as well as care provision more broadly. These operate within the broader societal context (C) of service users often being socially excluded and there being a high turnover of staff within these services and a high degree of variation across services in assessment tools that are recommended and in use, rendering the tool useful and relevant to the staff’s role.

The key actors (A) in implementing the SInQUE are the staff and service users, who require the skills and proactivity to administer the assessment and the motivation and trust to engage with the questions, respectively. Staff may encounter barriers such as a concern that some questions are too intrusive, and service users struggling more severely with their mental health may lack the concentration or motivation to engage with the questions. The potential outcomes operate through certain mechanisms (M), which include increased service user confidence and the prompting of more in-depth personal conversations between service users and staff. The tool also identifies more relevant priorities for service users, which may or may not be chosen as an active priority for support by staff owing to individual or organizational factors. Persistent and wide-ranging use of the tool could, over time, highlight the aspects of social inclusion that are feasible to work on and those that are regularly not being prioritized.

**Discussion**

**Principal Findings**

The online SInQUE was generally perceived as acceptable and potentially useful by supported accommodation staff and service users. This is consistent with findings of previous studies that used the SInQUE with other mental health populations [32,34]. Both staff and service users generally found the tool to be user-friendly and relevant and suggested that it could promote more targeted care planning and improve the relationship between staff and service users. Owing to the lower uptake of the SInQUE in residential care and floating outreach services, findings related to the tool’s utility in these settings are less conclusive than those for supported housing, where uptake was highest.

Some staff members expressed a concern that certain questions in the SInQUE could be perceived as intrusive by service users, indicating that they did not feel wholly comfortable asking what they perceived to be highly personal questions. However, this sentiment was not echoed by service users, who generally felt that the questions were appropriate and felt comfortable answering them. This finding is interesting given that supported...
accommodation service users have highlighted in previous research the importance of feeling personally understood by staff in their service and have endorsed a process of familiarization with staff [25].

We found that implementation support is essential to promote the use of the tool in services, as evidenced by the lack of use of the tool in the 2 regions where the SInQUE was introduced without a concerted implementation strategy. The most effective steps in our implementation strategy were those during which the use of the tool was actively endorsed by individuals in leadership positions, particularly service managers and local service leaders. However, even with our concerted implementation strategy, uptake of the SInQUE was only achieved with approximately 10% of service users living in mental health supported accommodation in the participating borough within the 5-month study period.

Limitations

We used an established, iterative process of testing and feedback to develop the online SInQUE and determine its real-world acceptability and utility for use in mental health supported accommodation. However, it is important to acknowledge certain limitations of this study.

As mentioned previously, uptake of the SInQUE tool was highest in supported housing compared with residential care homes and floating outreach support. It is unclear whether this discrepancy reflects a greater reluctance from staff or service users in residential care and floating outreach support to use the online SInQUE. As proposed by one residential care staff member, it is possible that the staff in these services perceived the tool as being less relevant to their role. The discrepancy in part reflects the greater number of 24-hour and 9-to-5 supported housing units in the borough compared with residential care and floating outreach services, with approximately 6 times as many service users living in supported housing compared with residential care and nearly twice as many living in supported housing compared with floating outreach support. Through the local health service community mental health rehabilitation team, we also had a more direct connection with supported housing teams compared with other service types, which may have further contributed to the imbalance in services in which the SInQUE was used.

There were no female service user participants in the qualitative analysis; therefore, the findings may not be applicable to women in supported accommodation. It is unclear why it proved more difficult to recruit female participants, although it may reflect the higher proportion of male service users availing of supported accommodation in England—one review suggests that between 68% and 74% of service users are male across all supported accommodation types [40]. Furthermore, as we only tested the SInQUE in 1 London borough, the findings may not be generalizable to other regions.

As the tool was only used with approximately 10% of service users in the borough, the success of our implementation strategy was limited, and the low uptake may limit the wider generalizability of our findings. Owing to the short period and limited scope of the study, it was also not possible to assess whether use of the SInQUE in practice led to improved outcomes for service users or how useful the repeat assessments were over time. As the staff who participated volunteered to do so and they chose which service users to complete the SInQUE with, the findings may have been affected by selection bias and may not accurately reflect all supported accommodation staff members’ and service users’ views.

Finally, we removed 1 question from the original SInQUE questionnaire for our online version as asking people whether they lived alone was considered redundant for people living in supported housing. We also made very minor changes to the wording of 2 other questions in response to users’ feedback (Table 2). We think it is unlikely that these modest changes substantially affected the SInQUE’s psychometric properties. However, revalidation of the SInQUE in its web-based form is desirable in the future to confirm its validity and determine whether the minor wording changes should be retained for all versions of the SInQUE.

Implications for Practice

The SInQUE can be recommended as a potentially useful and acceptable tool for use in mental health supported accommodation settings, particularly supported housing services that offer 24-hour or 9-to-5 support, to provide a thorough assessment of social inclusion and support care planning. The tool may help meet an identified wish from service users for more discussion and support with social inclusion and relationships [41]. It was evident during the study that there is currently no universal tool in widespread use to help with social inclusion in mental health supported accommodation, highlighting the potential gap for an assessment tool such as the SInQUE. If used widely across supported accommodation services, the online SInQUE has the potential to provide benchmarking data and identify service users’ most common priorities for greater social inclusion to inform service planning and evaluation.

Our findings also suggest that, for an assessment tool such as the SInQUE to be widely used, it is essential to have active leadership endorsement and support. For example, it may be required for managers or commissioners to direct staff to use the SInQUE with service users who are willing and reinforce this through team meetings, setting of use targets, or implementation of key performance indicators for its use.

Implications for Research

It is important to hear from staff and service users who chose not to use the online SInQUE to understand their reasons for not using the tool and highlight barriers to using the tool that we may have missed in this study. It would be useful to conduct further testing of the tool in residential care and floating outreach supported accommodation settings to better determine the utility of the SInQUE in these service types. It would also be useful to examine the utility of the SInQUE in other population groups within different service types to determine whether the tool may be useful in additional settings.

Future research is necessary to establish the level of uptake of the SInQUE that can be achieved in supported accommodation over a longer period and potentially establish more effective
A longer-term study is also needed to establish whether the possible benefits from using the SInQUE that were mentioned by staff and service users are achievable through the use of the tool and how any potential outcomes may vary over time. A hybrid implementation-evaluation study would address these queries to determine the effectiveness of the SInQUE tool as an intervention for social inclusion and establish a precise implementation strategy for widespread uptake of the tool in supported accommodation. Further research using the SInQUE is also warranted to examine service user needs related to social inclusion and identify any additional barriers to addressing these needs in supported accommodation services. Such research could be used to inform the development of a future complex intervention to support social inclusion in supported accommodation services.

Although this study chose to examine the utility of the online SInQUE specifically in supported accommodation, the tool may also be useful in other mental health populations. Previous studies have established that the SInQUE can be used with mental health service users with a wide range of diagnoses [13,32,34]. Therefore, it is reasonable to extrapolate that the online SInQUE may be useful to assess social inclusion and inform support and care planning for other mental health service users, not just those living in mental health supported accommodation.

Acknowledgments
This paper presents independent research funded by the National Institute for Health and Care Research School for Social Care Research. The views expressed are those of the authors and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

The authors would like to thank the supported accommodation managers, staff, and service users who supported and participated in this study. They would also like to thank the members of their expert advisory group and the University College London Service User Research Forum for their advice and feedback throughout the study.

Data Availability
The qualitative data generated and analyzed during this study are not publicly available to preserve participant anonymity. The quantitative data generated during this study are available from the corresponding author (BLE) upon reasonable request.

Authors’ Contributions
SE led recruitment, data collection, data analysis, and drafting of the paper. BLE led the study design and project management and supported the data analysis and drafting of the paper. HK, JC, PM, and GM co-designed the study and supervised the project. MD contributed to recruitment and data collection. GT contributed to the development of the study materials. All authors read, critically revised, and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Full description of the online Social Inclusion Questionnaire User Experience using the TIDieR (Template for Intervention Description and Replication) checklist for reporting interventions.
[DOCX File, 24 KB - ijmr_v13i1e45987_app1.docx ]

Multimedia Appendix 2
Laboratory testing topic guide.
[DOCX File, 17 KB - ijmr_v13i1e45987_app2.docx ]

Multimedia Appendix 3
Field-testing topic guides (for staff and service users).
[DOCX File, 16 KB - ijmr_v13i1e45987_app3.docx ]

Multimedia Appendix 4
Explanation of suggested changes during laboratory testing and field-testing that were not made after team review.
[DOCX File, 18 KB - ijmr_v13i1e45987_app4.docx ]

Multimedia Appendix 5
Interviews with staff and service users—illustrative quotes for each inductively derived subtheme.
References


35. SInQUE home page. SInQUE. URL: https://sinque.orwhiterose.ac.uk/84845/2/ScientificSummary_hsdr03050.pdf [accessed 2022-12-07]


Abbreviations

ICAMO: intervention-context-actor-mechanism-outcome
NHS: National Health Service
REC: Research Ethics Committee
SInQUE: Social Inclusion Questionnaire User Experience
TIDIE-R: Template for Intervention Description and Replication
UCL: University College London
A Web-Based Tool to Assess Social Inclusion and Support Care Planning in Mental Health Supported Accommodation: Development and Preliminary Test Study

URL: https://www.i-jmr.org/2024/1/e45987
doi:10.2196/45987
PMID:38477978

©Sharon Eager, Helen Killaspy, Joanna C, Gillian Mezey, Peter McPherson, Megan Downey, Georgina Thompson, Brynmor Lloyd-Evans. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 13.03.2024. 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.
Patterns of Skin Picking in Skin Picking Disorder: Ecological Momentary Assessment Study

Christina Gallinat1, PhD; Markus Moessner1, PD, PhD; Maximilian Wilhelm1,2, MSc; Nancy Keuthen3, PhD; Stephanie Bauer1,2,4, PD, PhD

1Center for Psychotherapy Research, University Hospital Heidelberg, Heidelberg, Germany
2German Center for Mental Health (DZPG), partner site Mannheim-Heidelberg-Ulm, Heidelberg, Germany
3Trichotillomania and Excoriation Disorder Program, Center for OCD and Related Disorders, Massachusetts General Hospital, Boston, MA, United States
4Heidelberg University, Heidelberg, Germany

Corresponding Author:
Christina Gallinat, PhD
Center for Psychotherapy Research
University Hospital Heidelberg
Bergheimer Straße 54
Heidelberg, 69115
Germany
Phone: 49 6221567876
Fax: 49 6221567350
Email: christina.gallinat@med.uni-heidelberg.de

Abstract

Background: Skin picking disorder (SPD) is an understudied mental illness that is classified as a body-focused repetitive behavior disorder. Literature suggests that pathological skin picking is strongly integrated into the daily lives of affected individuals and may involve a high degree of variability in terms of episode characteristics, frequency, and intensity. However, existing data on the phenomenology of SPD are limited and typically involve retrospective assessments, which may fail to accurately capture the behavior’s variability.

Objective: This study aimed to investigate skin picking in the daily lives of individuals with SPD by using ecological momentary assessment (EMA). The first aim focused on the description of skin picking patterns (eg, characteristics, intensity, and distribution of episodes and urges), and the second aim explored differences in characteristics and patterns between automatic and focused skin picking.

Methods: Participants were recruited online and underwent a web-based screening, a diagnostic telephone interview, and a comprehensive online self-report questionnaire before participating in an EMA protocol. The latter included 10 consecutive days with 7 pseudorandom, time-contingent assessments per day between 8 AM and 10 PM. The EMA questionnaire assessed the current skin picking urge, the occurrence of the behavior, and a detailed assessment of the episodes’ characteristics (eg, length, intensity, and consciousness) if applicable.

Results: The final sample consisted of 57 participants, who completed at least 70% of the scheduled assessments (n=54, 94.7% female: mean age 29.3, SD 6.77 years). They completed 3758 EMAs and reported 1467 skin picking episodes. Skin picking occurred frequently (mean 2.57, SD 1.12 episodes per day and person) in relatively short episodes (10-30 min; 10 min: n=642, 43.8%; 20 min: n=312, 21.3%; 30 min: n=217, 14.8%), and it was distributed quite evenly throughout the day and across different days of the week. Focused and automatic episodes were relatively balanced across all reported episodes (focused: n=806, 54.9%) and over the course of the day. The analyses showed statistically significant differences between self-reported triggers for the different styles. Visual or tactile cues and the desire to pick the skin were more important for the focused style (visual or tactile cues: mean focused style [Mf]=4.01, SD 0.69 vs mean automatic style [Ma]=3.47, SD 0.99; P<.001; SMD=0.64; desire to pick: Mf=2.61, SD 1.06 vs Ma=1.94, SD 1.03; P<.001; SMD=0.82), while boredom and concentration problems were more prominent in automatic skin picking (boredom: Mf=1.69, SD 0.89 vs Ma=1.84, SD 0.89; P=.03; SMD=–0.31; concentration problems: Mf=2.06, SD 0.87 vs Ma=2.31, SD 1.06; P=.006; SMD=–0.41).
Conclusions: These results contribute to an enhanced understanding of the phenomenology of SPD using a more rigorous assessment methodology. Our findings underscore that picking can impact affected persons multiple times throughout their daily lives.

Trial Registration: German Clinical Trials Register DRKS00025168; https://tinyurl.com/mr35pdwh

(KEYWORDS)
skin picking disorder; ecological momentary assessment; EMA; body-focused repetitive behavior; obsessive-compulsive spectrum; skin; dermatology; mental health; assessment; mobile phone

Introduction

Skin picking disorder (SPD) is a mental disorder, which is characterized by the body-focused repetitive behavior (BFRB) of manipulating one’s own skin including, for example, squeezing, scratching, or rubbing—summarized as “skin picking” [1]. With a lifetime prevalence of 1.4% to 3.1% [1,2], SPD is not a rare disorder, even though it received comparatively little attention in research and clinical practice so far.

Until now, there has been little research on the phenomenology of SPD, and the existing research is of questionable validity as it often entails retrospective reporting, so the clinical picture has not been described in sufficient detail to date. This hinders a well-grounded understanding of the disorder as well as the development of specific treatment options.

Few previous studies have described skin picking in terms of the frequency and episode length; for example, one study reported a median of 38 minutes for skin picking per day (range 1-360 min), while another found a mean of 8 (SD 22) episodes per day with an average length of 21 (SD 42) minutes [3,4]. In a more recent study, 78% of participants reported that they typically have 1 to 5 episodes per day and that most episodes are shorter than 30 minutes. Moreover, the majority reported that they picked their skin every day [5].

Meanwhile, data on high-risk times throughout the day are very scarce, with only 1 small study reporting such data (n=31) [6]. However, the small number of studies and the large variability among the results suggest a need for additional and more rigorous investigations.

In addition to episode characteristics, different styles of skin picking characterized by the extent of awareness during behavior were examined. “Focused skin picking” is hypothesized to occur more intentionally and in response to urges or difficult emotions, whereas “automatic skin picking” takes place without awareness and is supposed to be associated with certain (routine) situations and passive activities [7]. So far, little is known about the distribution of automatic and focused skin picking within and between individuals, other than that there seems to be high variability. However, a recent study reported a shift from focused skin picking toward more automatic skin picking with increasing age [8].

In terms of episode triggers, previous studies identified certain internal and external states commonly precipitating skin picking behavior. Commonly reported triggers are affective states (eg, tension or boredom), visual and tactile perceptions of skin irregularities, passive activities, and certain situations or places (eg, waiting, reading, or bathroom) [9-12]. Unfortunately, there is currently almost no data available on the distribution of skin picking and skin picking urges over the course of a day and a week.

Moreover, the existing studies on skin picking phenomenology include crucial shortcomings due to their cross-sectional and retrospective designs. It is well known that retrospective assessments imply a high risk of systematic biases, caused by the way memories are stored and retrieved [13]. Moreover, these designs are not able to capture dynamic processes and to identify specific variations, for example, in behavioral patterns throughout the day or week. Both of these issues are relevant to studies on SPD phenomenology. For example, the large range in the number and length of skin picking episodes in former studies indicates that it is critical to examine the distribution and characteristics of the behavior and to explore the role of intra-individual and inter-individual variability in the behavior.

In the clinical setting, affected individuals often report that the behavior can strongly vary from day to day—depending on a multitude of factors, for example, such as being in company versus alone or at work versus at home. These differences are masked in retrospective studies when the average time spent on skin picking in the last 2 weeks is assessed.

In addition, retrospective studies usually do not allow a reliable assessment and differentiation of characteristics of different styles of skin picking, which are characterized by the extent of awareness during skin picking. Moreover, the distribution of focused versus automatic skin picking as well as the link between specific triggers and different skin picking styles have not yet been investigated in detail. Of note, as most individuals with skin picking show a mixture of both styles, the retrospective assessment of separate triggers for automatic versus focused episodes would be very likely biased. However, the detailed investigation of skin picking styles and the associated triggers can serve as a solid basis for the specific selection and adaptation of interventions and behavioral strategies for certain risk situations or skin picking styles.

A promising method to comprehensively investigate processes of skin picking behavior is ecological momentary assessment (EMA) [14]. Momentary assessments within the daily life of individuals provide the opportunity to study dynamic processes in real time while minimizing retrospective biases. Since EMA allows a more detailed assessment of behavioral processes and implies a high ecological validity, the method received much attention in psychological research in the last 2 decades and was successfully applied by numerous studies in the...
investigation of different psychopathologies (eg, anxiety, substance use, or eating disorders) [15-17].

For skin picking research, EMA is a promising tool for reliably investigating the distribution as well as characteristics of skin picking episodes. The analysis of these data then affords an understanding of the course of skin picking behavior throughout the day and week in detail and identifies high-risk times and related circumstances. To the best of our knowledge, EMA has not yet been applied to investigate these research questions in SPD. Therefore, the main objective of this study was to investigate skin picking in the natural environment of individuals having SPD using EMA. Such data are urgently needed for a more comprehensive description and understanding of the phenomenology and mechanisms of this comparatively new disorder.

The study followed 2 aims: the first aim was to describe skin picking patterns in the daily lives of the participants (eg, number, length, intensity, distribution of skin picking episodes, distribution of skin picking urges, or self-reported triggers).

The second aim of this study was to explore differences between automatic and focused skin picking concerning distributions (eg, daytime), characteristics of the episodes (eg, length or intensity), and self-reported triggers.

**Methods**

**Procedures**

Participants were recruited between November 2021 and May 2022 through support groups and online via mailing lists, specific forums, and social media. Inclusion required a minimum age of 18 years; satisfaction of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria for SPD; and provision of informed consent for study participation. The inclusion of participants involved three stages of assessment: (1) an initial web-based screening, which assessed sociodemographic information and skin picking symptoms; (2) a diagnostic interview via telephone, in which the DSM-5 criteria for SPD were assessed; and (3) a web-based self-report questionnaire (baseline) for those assessed to be eligible in the interview.

EMA sampling started on the day after completion of the baseline questionnaire. The assessment period comprised 10 consecutive days with 7 pseudorandom, time-contingent assessments per day between 8 AM and 10 PM. In addition, participants were asked to record additional skin picking episodes (event contingent recording). The prompts were sent to the participants’ smartphones via text message, which contained a link to the EMA questionnaire. Additional records could be made via the web-based study platform. The time and event contingent EMA records took at most 5 minutes. All assessment procedures were conducted with the software ASMO [18].

**Measures**

**Screening**

The screening questionnaire included sociodemographic variables and the German version of the Skin Picking Scale-Revised (SPS-R) [19,20]. The scale assesses skin picking severity over the past week and consists of 8 items that can be split into 2 subscales: symptom severity and impairment. A global score can also be calculated. All items are rated on a 5-point Likert Scale from 0 (eg, “none”) to 4 (eg, “extreme”). The internal consistency of the total scale was high in this study (α=0.84; subscales: symptom severity: α=0.77 and impairment: α=0.85).

**Diagnostic Interview**

To assess the DSM-5 criteria for SPD, semistructured interviews based on a BFRB module (personal communication with L Mehrmann, February 2021) for the DIPS Open Access Diagnostic Interview for Mental Disorders were conducted via telephone [21]. The interviews were carried out by the first author (CG) and a student worker, who was trained and continuously supervised.

**Baseline Measures**

**Overview**

The baseline questionnaire contained the following assessment instruments.

**Skin Picking Severity**

The current skin picking severity was assessed in the baseline questionnaire with the SPS-R described above [19,20].

**Impairment due to Skin Picking**

Skin picking-related impairment was assessed with the German translation of the Skin Picking Impact Scale (SPIS) [22,23], which refers to the last week and contains 10 items capturing potential impairments due to skin picking (eg, feeling unattractive, ashamed, or not being able to do certain things due to skin picking) rated on a 5-point Likert scale (0: “not at all”; 4: “severe”). The internal consistency of the SPIS was excellent in this study (α=0.90).

**Modes of Skin Picking**

Different styles of skin picking (focused vs automatic) were assessed with the German version Milwaukee Inventory for the Dimensions of Adult Skin Picking (MIDAS) [7]. We translated the scale in a former study following generally accepted recommendations including backtranslation and approval by one of the authors of the original scale (DW Woods) [24]. The scale consists of 12 items, which are rated from 1 “not true for any of my skin picking” to 5 “true for all of my skin picking”. Both subscales (focused, automatic) contain 6 items and showed an acceptable internal consistency of α=0.62.

**Depressive Symptoms**

Depressive symptoms were captured using the Patient Health Questionnaire-9 (PHQ-9) [25]. The scale contains 9 items, which are rated on a Likert scale from 0 (“not at all”) to 3 (“almost every day”) in reference to the last 2 weeks. The scale showed a good internal consistency in our study (α=0.84).
Anxiety

Symptoms of generalized anxiety disorders were assessed with the Generalized Anxiety Disorders-7 (GAD-7) [26]. The Cronbach $\alpha$ was 0.84.

EMA Questionnaire

The EMA assessments included urge intensity (1: “no urge” to 5: “very strong”) and skin picking occurrence since the last assessment (yes or no). If skin picking occurred, additional questions assessed the following: intensity of skin picking (1: “very weak” to 5: “very strong”), length of the episode (12 options: about 10, 20, 30, …, 120 min), awareness at episode onset (“Did you notice when you started picking your skin?”; yes or no), and perceived triggers (“What contributed to your skin picking?”). For the last question, seven items had to be rated on a 5-point Likert scale: (1) visual or tactile cues, (2) itching, (3) tension, (4) boredom, (5) difficulties concentrating on a task, (6) desire for skin picking, (7) certain routine (e.g., evening routine), and (8) other (text field).

Statistical Analyses

Patterns of skin picking were analyzed using descriptive statistics. Frequencies for the number of episodes with certain characteristics (length, time of occurrence, or consciousness) were analyzed across all individuals and episodes. To control for the unequal number of skin picking episodes reported per person, mean scores within each person were calculated for urge intensity, episode intensity, and the rating of specific triggers. The average scores of the person means are reported. The distribution of skin picking urges as well as the distribution and characteristics of automatic and focused skin picking were also analyzed descriptively. t tests (2-tailed) for paired samples were calculated to test differences between focused and automatic episodes. Focused and automatic episodes were classified based on the yes or no question “Did you notice when you started picking your skin?” Differences were quantified using SMD. Analyses were performed in R (version 4.1.2; R Development Core Team, 2021) and with SPSS Statistics (version 29.0; IBM).

Ethical Considerations

All study procedures adhered to the latest version of the Declaration of Helsinki and were approved by the ethics committee of the Medical Faculty of Heidelberg University (S-222/2021). The trial was registered at the German Clinical Trials Register before recruitment started (DRKS00025168).

Participants provided informed consent for this study ahead of the initial screening and were able to discontinue participation at any time. Data are pseudonymized and can be subsequently matched to the respective persons only by authorized personnel. The data are anonymized as soon as possible after completion of the analyses.

All participants were provided with a €15 (approximately US $16) compensation in the form of a gift voucher for a bookstore. Additionally, if participants achieved an EMA completion rate of at least 70% (49 assessments), the voucher was upgraded to €50 (approximately US $54).

Results

Sample Description

Overall, 113 individuals completed the screening questionnaire. Of these, 79 (69.9%) participants started the EMA assessments. Further, 1 person dropped out during the EMA period and 1 person was excluded from the final analysis due to wearing an awareness bracelet, which vibrates when touching certain body areas for the prevention of skin picking. Overall, 57 (74%) out of 77 answered at least 70% of all scheduled EMA questionnaires (i.e., at least 49 assessments). The participant flow is shown in Figure 1.

Participants of the final sample (n=57) completed 65.93 (SD 7.24) EMAs on average, with a range of 51 to 99 per person. Frequencies above the number of scheduled time-contingent assessments (n=70) result from additional entries made by participants on their own initiative (event-contingent records).
Participants

The majority of participants were female (54/57, 94.7%) with a mean age of 29.3 (SD 6.77) years. About half (28/57, 49.1%) of the participants were employed and one-third (18/57, 31.6%) were university students. The sample showed a PHQ-9 mean score of 11.63 (SD 5.41), indicating moderate depressive symptoms; a GAD-7 mean score of 9.63 (SD 4.85), indicating mild to moderate anxiety; and a mean SPS-R score of 18.00 (SD 4.00), indicating substantial SPD severity. The participants in the analyzed sample do not differ from the participants who were excluded from the analyses due to the low EMA completion rate (less than 49 assessments, <70%). t tests (2-tailed) for independent samples and χ²-quadrat tests did not yield any statistically significant differences in terms of the assessed sociodemographic and clinical variables (all *P* > .05).

A detailed overview of the sample characteristics is given in Table 1.
Table 1. Sample characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total EMA&lt;sup&gt;a&lt;/sup&gt; sample (n=77)</th>
<th>EMA≥70% (n=57)</th>
<th>EMA&lt;70% (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex, n (%)</td>
<td>74 (96.1)</td>
<td>54 (94.7)</td>
<td>20 (100)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>28.84 (6.51)</td>
<td>29.3 (6.77)</td>
<td>27.3 (5.74)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Still in school</td>
<td>1 (1.3)</td>
<td>—</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Middle secondary</td>
<td>8 (10.4)</td>
<td>6 (10.5)</td>
<td>2 (10)</td>
</tr>
<tr>
<td>Highest secondary</td>
<td>27 (35.1)</td>
<td>19 (33.3)</td>
<td>8 (40)</td>
</tr>
<tr>
<td>University</td>
<td>41 (53.2)</td>
<td>32 (56.1)</td>
<td>9 (45)</td>
</tr>
<tr>
<td>Occupational status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>40 (51.9)</td>
<td>28 (49.1)</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Trainee</td>
<td>1 (1.3)</td>
<td>1 (1.8)</td>
<td>—</td>
</tr>
<tr>
<td>School student</td>
<td>1 (1.3)</td>
<td>—</td>
<td>1 (5)</td>
</tr>
<tr>
<td>University student</td>
<td>23 (29.9)</td>
<td>18 (31.6)</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Housewife or househusband</td>
<td>3 (3.9)</td>
<td>3 (5.3)</td>
<td>—</td>
</tr>
<tr>
<td>Retired</td>
<td>2 (2.6)</td>
<td>1 (1.8)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>2 (2.6)</td>
<td>1 (1.8)</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (6.5)</td>
<td>5 (8.8)</td>
<td>—</td>
</tr>
<tr>
<td>Family status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single, n (%)</td>
<td>34 (44.2)</td>
<td>22 (38.6)</td>
<td>12 (60)</td>
</tr>
<tr>
<td>In a relationship, n (%)</td>
<td>25 (32.5)</td>
<td>20 (35.1)</td>
<td>5 (25)</td>
</tr>
<tr>
<td>Married, n (%)</td>
<td>15 (19.5)</td>
<td>12 (21.1)</td>
<td>3 (15)</td>
</tr>
<tr>
<td>Separated or divorced, n (%)</td>
<td>2 (2.6)</td>
<td>2 (3.5)</td>
<td>—</td>
</tr>
<tr>
<td>Other, n (%)</td>
<td>1 (1.3)</td>
<td>1 (1.8)</td>
<td>—</td>
</tr>
<tr>
<td>PHQ-9&lt;sup&gt;c&lt;/sup&gt;, mean (SD)</td>
<td>11.95 (5.6)</td>
<td>11.63 (5.41)</td>
<td>12.15 (5.86)</td>
</tr>
<tr>
<td>GAD-7&lt;sup&gt;d&lt;/sup&gt;, mean (SD)</td>
<td>10.04 (4.64)</td>
<td>9.63 (4.85)</td>
<td>10.5 (4.01)</td>
</tr>
<tr>
<td>SPS-R&lt;sup&gt;e&lt;/sup&gt;, mean (SD)</td>
<td>17.69 (3.98)</td>
<td>18 (4)</td>
<td>16.7 (3.87)</td>
</tr>
<tr>
<td>SPIS&lt;sup&gt;f&lt;/sup&gt;, mean (SD)</td>
<td>23.45 (8.57)</td>
<td>23.33 (8.94)</td>
<td>23.8 (7.61)</td>
</tr>
<tr>
<td>MIDAS focused&lt;sup&gt;g&lt;/sup&gt;, mean (SD)</td>
<td>19.9 (3.92)</td>
<td>19.7 (3.9)</td>
<td>20.45 (4.05)</td>
</tr>
<tr>
<td>MIDAS automatic&lt;sup&gt;h&lt;/sup&gt;, mean (SD)</td>
<td>18.18 (3.63)</td>
<td>18.49 (3.58)</td>
<td>17.3 (3.74)</td>
</tr>
</tbody>
</table>

<sup>a</sup>EMA: ecological momentary assessment.
<sup>b</sup>Not available.
<sup>c</sup>PHQ-9: Patient Health Questionnaire-9, depressive symptoms.
<sup>d</sup>GAD-7: generalized anxiety disorders-7.
<sup>e</sup>SPS-R: Skin Picking Scale-Revised.
<sup>f</sup>SPIS: Skin Picking Impact Scale.
<sup>g</sup>MIDAS focused: Milwaukee Inventory for the Dimensions of Adult Skin Picking, focused skin picking.
<sup>h</sup>MIDAS automatic: Milwaukee Inventory for the Dimensions of Adult Skin Picking, automatic or unconscious skin picking.

**Number and Distribution of Episodes**

In total, 57 participants completed 3758 EMAs and reported 1467 skin picking episodes during the EMA period of 10 days. Altogether, 1351 (92.1%) episodes were reported in time-based assessments and only 116 (7.9%) in event-based assessments. On average, participants reported a mean number of 2.57 (SD 1.12; range 0.8-5.4) episodes per day.

Slightly more than half of the sample (32/57, 56.1%) reported episodes on each day of the 10-day EMA phase, while 28.1% (n=16) reported 1 day without skin picking and 15.8% (n=9) had 2, 3, or 4 days without skin picking.

Skin picking episodes were relatively evenly distributed throughout the day. Small peaks in the number of episodes emerged in the first (8-10 AM; n\textsubscript{episodes}=253, 17.3% of all 1467 episodes) and the last (8-10 PM; n\textsubscript{episodes}=211, 14.4% of all 1467 episodes) regular assessment period of each day. The number of episodes over the course of a day is shown in detail in Figure 2 and Table 2. It should be noted that to avoid a biased comparison between time-based and event-based surveys, only the periods covered by the time-based assessment are presented.

Figure 2. Episode distribution throughout the day (event- and time-based EMAs; 8 AM-10 PM). EMA: ecological momentary assessment.

Table 2. Urge and episode parameters over the day\textsuperscript{a}.

<table>
<thead>
<tr>
<th>Time</th>
<th>All episodes (N=1467), n (%)</th>
<th>Focused episodes (n=806), n (%)</th>
<th>Automatic episodes (n=661), n (%)</th>
<th>Intensity of episodes (n=1467), mean (SD)</th>
<th>Urge intensity EMA\textsuperscript{b} with episodes (n=1467), mean (SD)</th>
<th>Urge intensity EMA without episodes (n=2291), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-10 AM</td>
<td>253 (17.2)</td>
<td>156 (19.35)</td>
<td>97 (14.67)</td>
<td>2.61 (0.87)</td>
<td>2.53 (0.98)</td>
<td>1.78 (0.79)</td>
</tr>
<tr>
<td>10 AM to Noon</td>
<td>202 (13.8)</td>
<td>108 (13.4)</td>
<td>94 (14.22)</td>
<td>2.38 (0.72)</td>
<td>2.54 (0.92)</td>
<td>1.92 (0.73)</td>
</tr>
<tr>
<td>Noon to 2 PM</td>
<td>201 (13.7)</td>
<td>103 (12.78)</td>
<td>98 (14.83)</td>
<td>2.19 (0.78)</td>
<td>2.75 (0.86)</td>
<td>1.87 (0.68)</td>
</tr>
<tr>
<td>2-4 PM</td>
<td>173 (11.8)</td>
<td>86 (10.67)</td>
<td>87 (13.16)</td>
<td>2.43 (0.86)</td>
<td>2.89 (0.84)</td>
<td>1.83 (0.71)</td>
</tr>
<tr>
<td>4-6 PM</td>
<td>163 (11.1)</td>
<td>85 (10.55)</td>
<td>78 (11.8)</td>
<td>2.43 (0.88)</td>
<td>2.90 (0.95)</td>
<td>1.92 (0.76)</td>
</tr>
<tr>
<td>6-8 PM</td>
<td>180 (12.3)</td>
<td>99 (12.28)</td>
<td>81 (12.25)</td>
<td>2.43 (0.78)</td>
<td>3.02 (0.88)</td>
<td>1.92 (0.71)</td>
</tr>
<tr>
<td>8-10 PM</td>
<td>211 (14.4)</td>
<td>108 (13.4)</td>
<td>103 (15.58)</td>
<td>2.77 (0.74)</td>
<td>3.02 (0.88)</td>
<td>1.94 (0.76)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Only periods of time-based assessments are listed. The total N refers to all registered episodes (N\textsubscript{episodes}=1467, N\textsubscript{persons}=57). Intensity scores are average scores of person-wise means.

\textsuperscript{b}EMA: ecological momentary assessment.

Weekdays
The episodes were quite evenly distributed over the days of the week. Across all participants, the average number of episodes per day ranged between 2.20 for Saturdays and 2.77 for Mondays and Tuesdays (Monday: mean 2.77, SD 1.78; Tuesday: mean 2.77, SD 1.65; Wednesday: mean 2.66, SD 1.50; Thursday: mean 2.54, SD 1.50; Friday: mean 2.53, SD 1.59; Saturday: mean 2.20, SD 1.50; and Sunday: mean 2.57, SD 1.66).

Length and Intensity
Of all 1467 episodes, participants indicated the shortest selectable length (approximately 10 minutes) in 43.8% (n=642), 20 minutes in 21.3% (n=312), and 30 minutes in 14.8% (n=217; Table 3). Only 9 (15.8%) participants reported any episode of 60 minutes or longer and only 6 (10.5%) reported episodes of at least 90 minutes.

Table 2 displays the distribution of focused and automatic episodes as well as episode intensity and urge intensity over the course of the day.

The reported intensity of the episodes across all subjects was on average 2.55 (SD 1.11; 2: “mild”, 3: “medium”). Throughout the day, the intensity of the episodes was quite stable. The average person means in the regular EMA phase (8 AM-10 PM) varied between 2.19 (SD 0.78; noon to 2 PM) and 2.77 (SD 0.74; 8-10 PM). Slightly higher average intensities were reported in the evening and the morning (see Table 2).

Overall, in terms of intensity, most episodes were rated as very mild (288/1467, 19.6%), mild (n=448, 30.5%), or medium (n=426, 29%). Participants rated 16.4% (n=240) of the episodes as severe and 4.4% (n=65) as very severe. Further, 10 (17.5%) participants did neither report severe nor very severe episodes.
Table 3. Length of episodes.

<table>
<thead>
<tr>
<th>Approximate length</th>
<th>Episodes, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 min</td>
<td>642 (43.8)</td>
</tr>
<tr>
<td>20 min</td>
<td>312 (21.3)</td>
</tr>
<tr>
<td>30 min</td>
<td>217 (14.8)</td>
</tr>
<tr>
<td>40 min</td>
<td>97 (6.6)</td>
</tr>
<tr>
<td>50 min</td>
<td>39 (2.7)</td>
</tr>
<tr>
<td>60 min</td>
<td>64 (4.4)</td>
</tr>
<tr>
<td>70 min</td>
<td>14 (0.9)</td>
</tr>
<tr>
<td>80 min</td>
<td>26 (1.8)</td>
</tr>
<tr>
<td>90 min</td>
<td>22 (1.4)</td>
</tr>
<tr>
<td>100 min</td>
<td>6 (0.4)</td>
</tr>
<tr>
<td>110 min</td>
<td>2 (0.1)</td>
</tr>
<tr>
<td>120 min</td>
<td>26 (1.8)</td>
</tr>
</tbody>
</table>

aAll episodes reported in time- and event-based ecological momentary assessments (N<sub>episodes</sub>=1467; N<sub>persons</sub>=57).

**Urge Intensity**

The mean urge intensity (average scores of person-wise means) in assessments with reported episodes varied between 2.53 (SD 0.98) in the morning (8-10 AM) and increased in small increments throughout the day with the highest mean being 3.02 (SD 0.88) in the evening (8-10 PM). So, the average urge intensity varied between mild (“2”) and medium (“3”) and was significantly higher in assessments with reported episodes (mean 2.84, SD 0.71) compared to those without episodes (mean 1.89, SD 0.65; t<sub>56</sub>=12.31, P<.001; SMD=1.63). The average scores for the urge intensity per period are shown in Table 2.

**Episode Characteristics**

**Consciousness**

Participants reported a conscious onset of the behavior in 54.9% (n=806; “focused episodes”) and an unconscious onset in 45.1% (n=661; “automatic episodes”) of all 1467 episodes. Most participants reported both types of episodes (49/57, 86%). One-third of participants (n=19) reported 75% or more focused episodes and 8 (14%) patients of these reported exclusively focused episodes. A proportion of 75% or more automatic episodes was indicated by 8 (14%) participants, and overall, it ranged between 0% and 98.2% (median 39.3, IQR 14.2-63.3).

Across all participants, the ratio between these 2 modes was relatively balanced throughout the day, with focused episodes occurring slightly more often. However, comparatively more focused episodes occurred in the morning (8-10 AM). Details are shown in Table 2.

Focused and automatic episodes did not differ significantly in terms of the intensity of the behavior or urge intensity (intensity: mean focused style [M<sub>f</sub>]=2.56, SD 0.62; mean automatic style [M<sub>a</sub>]=2.45, SD 0.78; t<sub>48</sub>=1.52, P=.14; urge intensity: M<sub>f</sub>=2.86, SD 0.77; M<sub>a</sub>=2.90, SD 0.86; t<sub>48</sub>=–0.19, P=.85).

**Self-Reported Triggers**

Across all participants, the highest average values resulted for visual or tactile cues (eg, felt or seen something on the skin; mean 3.64, SD 1.26), tension (mean 2.63, SD 1.29), and habit (mean 2.71, SD 1.45).

Comparisons between focused and automatic episodes showed higher scores in focused episodes for visual or tactile cues as well as for the item “wanted to pick the skin” (SMD=0.64 and 0.82, respectively). In contrast, boredom and problems with concentration achieved higher scores in automatic episodes (SMD=–0.31 and –0.41, respectively).

In the “other” category, additional conditions were mentioned in 97 episodes: working or being at the PC, talking on the phone, smartphone time, reading, watching television, driving, showering, encountering a mirror, physical fatigue or tiredness, hunger, emotional discomfort, and social situations or conflicts. Scores are displayed in Table 4.
Principal Findings

SPD has now been officially recognized as a separate disorder for more than 10 years. However, despite increased research efforts, there is still a lack of studies on the phenomenology of the disorder. To our knowledge, this is the first study to investigate skin picking behavior by using EMA in the daily life of people with SPD.

The results document in several ways how strongly the behavior is interwoven with the everyday life of affected individuals. For example, 56.1% (32/57) reported that they experienced no day without skin picking within the 10-day study phase, but only 15.8% (n=9) reported 2 to 4 days without skin picking. In other words, skin picking occurred almost every day. In addition, participants reported an average of 2.6 episodes per day (range 0.8-5.4), suggesting that the behavior is not limited to 1 daily episode, but occurs several times a day and continuously influences daily life. These results are consistent with the results of 2 retrospective studies reporting also several episodes per day [3,5]. The continuity of the behavior is also reflected by the results over the course of the day and the week. Throughout the day, episodes were more or less evenly distributed, with only small peaks in the morning and evening. Similarly, the average urge intensity varied only slightly over the monitored periods and ranged constantly between weak and medium, with values in the evening being somewhat higher. However, as expected, the urge intensity was considerably higher in assessments with reported episodes compared to those without.

Regarding the frequency of the episodes, there were also only small differences between the different weekdays. The lowest average number of episodes was reported for Saturdays and the highest for Mondays and Tuesdays, but the differences between other weekdays were quite small. Overall, data regarding the skin picking urges and behavior indicate that both are experienced frequently by affected individuals.

In terms of the episode characteristics, it is important to note that 43.8% (n\textsubscript{episodes}=642/1467) were no more than 10 minutes long and 80% (n\textsubscript{episodes}=1171) of the episodes were no longer than 30 minutes, so the results suggest rather short, but frequent episodes. This is also in line with previous studies reporting that the majority of episodes are under 30 minutes [3,5]. However, short episodes are not necessarily mild since the skin can be severely damaged in just a few minutes.

Regarding consciousness of the episodes, the results show groups of individuals with a quite high preponderance (eg, \geq 75\% of episodes) of a focused (19/57, 33\%) or automatic (8/57, 14\%) style. A unilateral skin picking style, where individuals predominantly (>95\% of all episodes) show either automatic or focused skin picking, was relatively rare (automatic: 2/57, 4\%; focused: 8/57, 14\% of the sample).

However, the ratio between focused and automatic episodes was relatively balanced, although there were clear differences between individuals. Overall, more participants showed a tendency toward a focused style. The minor predominance of focused skin picking is also consistent with the results of a recent study that similarly found a slight dominance of focused skin picking for middle adulthood [8].

In recent years, different studies tried to identify different skin picking subtypes between individuals regarding various characteristics (eg, symptom presentation and styles of skin picking, or neurobiology), but nevertheless, this research is still in its beginning [7,27-30]. However, as research shows that most people with SPD show both styles of skin picking, there is an obvious necessity to understand the different types of pathological skin picking to develop prevention and intervention strategies specifically for automatic and focused skin picking. This is especially the case because the onset and course of an automatic episode can strongly differ from focused episodes necessitating different coping strategies matched to the specific picking style.

The results showed statistically significant differences between self-reported triggers for automatic and focused episodes: visual or tactile cues and the desire to pick the skin (item “wanted to pick”) played a more important role in focused episodes, while boredom and problems with concentration were more related to automatic episodes. Other triggers (eg, tension or itch) did not differ between the 2 modes of skin picking. The largest

### Table 4. Self-reported triggersa.

<table>
<thead>
<tr>
<th>Trigger</th>
<th>Total (N=57), mean (SD)</th>
<th>Focused (n=49), mean (SD)</th>
<th>Automatic (n=49), mean (SD)</th>
<th>t testb (df)</th>
<th>P value</th>
<th>SMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual or tactile cues</td>
<td>3.64 (1.26)</td>
<td>4.01 (0.69)</td>
<td>3.47 (0.99)</td>
<td>4.482 (48)</td>
<td>&lt;.001</td>
<td>0.64</td>
</tr>
<tr>
<td>Tension</td>
<td>2.63 (1.29)</td>
<td>2.67 (0.89)</td>
<td>2.84 (0.97)</td>
<td>-1.532 (48)</td>
<td>.13</td>
<td>-0.22</td>
</tr>
<tr>
<td>Boredom</td>
<td>1.69 (1.03)</td>
<td>1.69 (0.89)</td>
<td>1.84 (0.89)</td>
<td>-2.187 (48)</td>
<td>.03</td>
<td>-0.31</td>
</tr>
<tr>
<td>Problems with concentration</td>
<td>2.17 (1.31)</td>
<td>2.06 (0.87)</td>
<td>2.31 (1.06)</td>
<td>-2.847 (48)</td>
<td>.006</td>
<td>-0.41</td>
</tr>
<tr>
<td>Wanted to pick the skin</td>
<td>2.14 (1.27)</td>
<td>2.61 (1.06)</td>
<td>1.94 (1.03)</td>
<td>5.753 (48)</td>
<td>&lt;.001</td>
<td>0.82</td>
</tr>
<tr>
<td>Habit or routine</td>
<td>2.71 (1.45)</td>
<td>2.82 (1.07)</td>
<td>2.56 (1.10)</td>
<td>1.818 (48)</td>
<td>.08</td>
<td>0.26</td>
</tr>
<tr>
<td>Itch</td>
<td>1.67 (1.09)</td>
<td>1.75 (0.94)</td>
<td>1.81 (1)</td>
<td>-0.594 (39)</td>
<td>.556</td>
<td>-0.09</td>
</tr>
</tbody>
</table>

aAnswers rated on a 5-point Likert scale (1: not at all; 5: extremely). t test results refer to comparisons of the average scores of person-means in focused and automatic episodes (t\textsubscript{episodes}=1295, n\textsubscript{person}=49). Further, 8 persons were excluded from the comparison as they reported no automatic episodes.
b2-tailed.
difference was found for the trigger desire (“wanted to pick”; SMD=0.82). Of note, the results do not provide any evidence that 1 of the 2 styles is associated more strongly with tension than the other.

**Strengths and Limitations**

Overall, the results offer useful insights into the nature, frequency, distribution, and intensity as well as specific triggers of skin picking. They also provide important starting points for future studies that should investigate these aspects in more detail. However, our results should be interpreted in light of the specific strengths and limitations of this study. The latter may include a bias due to the self-selection of the participants. It is likely a rather specific sample of individuals, who are willing to track their skin picking for a period of 10 days several times a day. However, our data suggest a substantial impairment in terms of skin picking severity, depression, anxiety, and skin picking–related impairment.

Another limitation results from the assessment method since self-observation and tracking skin picking can also increase the awareness and therefore the controllability of the behavior. Moreover, it is also discussed that the registration of an episode may serve a punishing function due to the extra effort to record it so that the likelihood of the behavior is reduced. These mechanisms could have produced 2 biases in this study: first, the number of automatic episodes could be underestimated due to the increased awareness during this study. Second, the frequency and intensity of the behavior may have been reduced by the continuous monitoring within this study’s period.

Also, the assessment started regularly with the question “Have you picked your skin since the last assessment?” This could have caused a bias toward more reported episodes in the first period of the day as individuals might also report skin picking, which occurred in the night before. Consequently, the total number of the period between 8 and 10 AM should be interpreted cautiously.

Another limitation refers to the assessment of the episode length, which was assessed by multiple choice with options in steps of 10 minutes. The shortest selectable option was “about 10 minutes,” but during this study, we received feedback from participants that their episodes were much shorter. However, this also leads us to the open question of what constitutes a skin picking episode and if micro episodes might play an important role. In addition, we know from clinical work that some people report that the behavior occurs almost constantly throughout the day. In this context, the question arises, whether and for whom it makes sense to divide the behavior into episodes. In this study, participants were forced to report behavioral episodes, but it remains unclear what participants have defined as an episode for themselves and if they tracked microepisodes. Future research needs to address these issues by applying an even tighter, more precise measurement of behavior, but this will also need to take the aforementioned difficulty of measurement reactivity into account.

Despite these limitations and the need for further research, this study also has some important strengths. To the best of our knowledge, this is the first study using EMA to assess skin picking, and it is also the first EMA study in the field of pathological BFRBs in general. It provides new insights into the phenomenology of the SPD allowing for a more reliable and accurate description of skin picking in the everyday life of affected individuals, which is essential for a comprehensive understanding of SPD of this relatively newly defined disorder. The study clearly demonstrates the advantages of measurement via EMA, since behavioral parameters could be assessed that cannot be measured at all—or only with considerable distortions—in retrospective assessments. Furthermore, this study was conducted with a sample of individuals fulfilling the diagnostic criteria for SPD, who showed good adherence overall, so this study provides high-quality data allowing for a detailed analysis of the phenomenology of SPD.

Our experience with the assessment of skin picking using EMA and the resulting data serve as a firm basis for further EMA studies on SPD and other disorders in the field of BFRBs and contribute to an enhanced understanding of an understudied but highly impairing mental disorder.

**Acknowledgments**

We would like to thank all participants, the BFRB community, and the TLC Foundation for BFRBs. The study would not have been possible without your support. We also thank Nadine Treß, MSc, who contributed to this study by conducting diagnostic interviews. This study was funded by the TLC Foundation for BFRBs as part of the 2020 Early Career Research grant. Generative artificial intelligence was not used in this study in any aspect of this paper’s writing.

**Data Availability**

The data sets analyzed during this study are not publicly available to protect participant privacy.

**Authors' Contributions**

CG handled the conceptualization, methodology, data curation, formal analysis, writing of the original draft, project administration, and funding acquisition. MM acted on the conceptualization, methodology, supervision, and review and editing of the writing. MW worked on the conceptualization, methodology, visualization, and review and editing of the writing. NK did the supervision, and review and editing of the writing. SB dealt with the conceptualization, methodology, resources, supervision, and review and editing of the writing.
Conflicts of Interest
None declared.

References


Abbreviations

- **BFRB**: body-focused repetitive behavior
- **DSM-5**: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
- **EMA**: ecological momentary assessment
- **GAD-7**: Generalized Anxiety Disorders-7
- **M_1**: mean automatic style (MIDAS)
- **M_2**: mean focused style (MIDAS)
- **MIDAS**: Milwaukee Inventory for the Dimensions of Adult Skin Picking
- **PHQ-9**: Patient Health Questionnaire-9
- **SPD**: skin picking disorder
- **SPIS**: Skin Picking Impact Scale
- **SPS-R**: Skin Picking Scale-Revised

©Christina Gallinat, Markus Moessner, Maximilian Wilhelm, Nancy Keuthen, Stephanie Bauer. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.07.2024. 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.
Effects of Background Music on Attentional Networks of Children With and Without Attention Deficit/Hyperactivity Disorder: Case Control Experimental Study

Camila Guimarães Mendes¹, PhD; Jonas Jardim de Paula², PhD; Débora Marques Miranda¹, PhD

¹Department of Pediatrics, Federal University of Minas Gerais, Belo Horizonte, Brazil
²Department of Mental Health, Federal University of Minas Gerais, Belo Horizonte, Brazil

Corresponding Author:
Camila Guimarães Mendes, PhD
Department of Pediatrics
Federal University of Minas Gerais
Avenida Professor Alfredo Balena, 190 sl85
Belo Horizonte, 30130100
Brazil
Phone: 55 3134099663
Email: camilagmbh@gmail.com

Abstract

Background: To sustain performance during a task that requires attention may be a challenge for children with attention deficit/hyperactivity disorder (ADHD), which strongly influences motivation for tasks and has been connected to the level of arousal.

Objective: This study aimed to analyze the effect of musical stimulus on attentional performance in children with ADHD and typically developing children.

Methods: A total of 76 boys (34 with ADHD and 42 typically developing) performed the Attention Network Test (ANT) for children under 2 experimental conditions (with and without music). Four attentional measures were extracted from the ANT. We tested the effect of the experimental condition and its interaction with the group using repeated measures ANOVA.

Results: We found no significant main effects or interactions for the reaction times of the alerting, orienting, and conflict attentional networks of the ANT (all \( P > .05 \)). Regarding ANT errors, we found a significant main effect for music, with a moderate effect size (\( F_{1,72} = 9.83; P = .03; \eta^2 = 0.06 \)) but the condition \times group interaction was not significant (\( F_{1,72} = 1.79; P = .18 \)). Participants made fewer errors when listening to music compared to the control condition.

Conclusions: Music seems not to interfere in the attentional network in children and adolescents. Perhaps background music affects motivation. Future studies will be needed to validate this.

Trial Registration: ReBEC.gov U1111-12589039; https://ensaiosclinicos.gov.br/rg/RBR-8s22sh8

KEYWORDS
attention; background music; ADHD; children; adolescents; music; attention network; effects; preliminary study; attention deficit/hyperactivity disorder

Introduction

Attention deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder characterized by harmful levels of inattention, impulsivity, and hyperactivity [1]. ADHD exhibits considerable heterogeneity, with individuals’ symptoms reflecting impairments in different cognitive aspects [2], causing distress or problems at home, at school, and with peers [3]. Impaired cognitive aspects in ADHD include frequent compromises in executive functions (ie, working memory, inhibitory control, cognitive flexibility, planning, and problem-solving), self-regulation states (ie, the purposeful or automatic mechanisms that enable behavior to be adapted appropriately to a changing context), motivation (ie, temporal reward discounting), and time perception (ie, the ability to discriminate and compare time intervals) [4-6].
hyperactivity and inattention levels of children with ADHD are noticeably higher than expected.

The attentional modeling of Berger and Posner [7] provides an appropriate theoretical framework to account for ADHD dysfunction because it conceptualizes most of the abilities mentioned above as part of attentional networks, such as alerting (ie, arousal of the cognitive system), orienting (ie, allocating attentional focus in the visual field), and executive control (ie, ability to control our own behavior, resolve conflict, and inhibit impulsive responses). A task that requires extra effort for children to sustain performance may be a challenge for children with ADHD, especially in suboptimal conditions [8]. On the other hand, effort is determined by the motivation to perform the task and has been connected to the level of arousal and activation [8,9]. This explains why children with ADHD, who are easily distracted by external stimuli, may benefit from stimuli that promote increased alertness and consequently improve performance in the task [10,11].

A recent systematic review showed listening to music without lyrics that was chosen by the listener seemed to improve performance in tasks requiring attention [12]. Music enhances arousal, can affect mood, and increases motivation, especially when it is preferred by the listener, potentially benefiting the learning process through emotional processes [13-15]. This heightened state of alertness and pleasant mood can enhance attentional resources, allowing the listener to concentrate better and sustain focus on cognitive tasks [16,17]. Music holds the potential to augment the emotion regulation abilities and mood of young individuals in their daily experiences [18].

Knowledge of the effect of music on the cognitive function of individuals with ADHD is still limited due to inconsistent results [19]. Among studies that evaluated music as a form of stimulation in ADHD, 2 reported improvements in mathematical problem-solving [20,21], while another study assessing schoolwork completion (including math, reading, reading comprehension, and language arts) showed no significant difference in cognitive function. The heterogeneity in the methodology of these studies makes it difficult to draw conclusions on the true effect of music on task performance. Nevertheless, a recent review indicated that listening to music can reduce symptoms of ADHD and improve timing perception and regulation [22], which are important for the functionality and well-being of this population.

At present, there is a lack of data assessing the impact of music listening on the attention networks of children with ADHD. Therefore, the aim of this study was to investigate the effects of music listening on the attention networks—namely, alerting, orienting, and conflict—in children with ADHD and typically developing children, while also exploring the relationship with the attentional profile of these children. Given that previous studies involving ADHD [23,24] incorporated measures of error types alongside conventional assessments of the 3 attention networks, we will also examine whether music influences error rates during task performance. Our hypothesis is that music may enhance attentional performance in children with ADHD differently from their typically developing peers.

**Methods**

**Study Design**

This preliminary, experimental, repeated-measures study was conducted from 2019 to 2022 to explore the impact of music listening compared to no music listening on attention performance. We enrolled boys aged 10 to 12 years, both with and without ADHD, who completed the Attention Network Test (ANT) for children twice under randomized conditions.

**Recruitment**

A total of 76 boys aged 10 to 12 years participated, comprising 34 with ADHD and 42 without ADHD (Figure 1). This age range was selected based on evidence indicating that children younger than 10 years are still developing their musical preferences, while adolescents tend to be more receptive to unfamiliar music styles [25]. Given that the musical stimulus in our study needed to be familiar and preferred by the listener, we focused on the age range of 10 to 12 years. Also, only boys were included, because the majority of children treated at the university hospital were male.
Figure 1. Flow diagram of the research process. ADHD: attention deficit/hyperactivity disorder; TD: typically developing.

Participants were recruited from a university hospital that provides psychiatric care for children and adolescents with ADHD, as well as pediatric follow-up for healthy individuals. Children with ADHD met the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, criteria [1] and underwent assessment using the semistructured Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime (K-SADS-PL) [26]. Moreover, they achieved scores at or above the 10th percentile on the Brazilian version of Raven’s Colored Progressive Matrices intelligence test [27].

The control group was selected from the local community between 2021 and 2022. This group was matched with the ADHD group in terms of age and socioeconomic status, and they met the inclusion criteria by not having a diagnosis of ADHD or by not scoring above the cutoff points on screening questionnaires for ADHD. These cutoff points included having more than 5 ADHD symptoms identified by the Swanson, Noland, and Pelham Scale IV (SNAP IV) or having a t score greater than 70 on the ADHD scale of the Child Behavior Checklist for Ages 6-18 (CBCL/6-18) [28,29].

**Intervention**

The intervention required participants to perform an attention task in 2 different conditions: with music and without music. The music selection comprised 5 songs chosen through interviews with children aged 10 to 12 years, who shared their favorite and most frequently listened-to songs. These songs were played during the test. It is important to note that the children interviewed about their favorite music were not necessarily participants in the study.

To gauge the emotional connection between listeners and songs (including familiarity, preference, mood, and arousal), a questionnaire was administered. Participants listened to song excerpts and answered questions such as “Do you know this song?” (answers were yes, maybe, or no), “Do you like this song?” (answers were yes, neutral, or no), and “How do you feel listening to these songs?” using the adapted Self-Assessment Manikin Scale [30]. A 5-point Likert scale was used to rate subjective mood (1=very sad, 2=sad, 3=neutral, 4=happy, and 5=very happy) and arousal (1=nonarousal, 2=low arousal, 3=neutral, 4=arousal, and 5=high arousal) based on images pointed to by the children. This questionnaire was administered before the ANT to ensure that the results were not influenced by the child’s performance. It can be found in Multimedia Appendix 1.

To prevent experimenter bias, the order of play of the songs was determined through a random drawing using Microsoft Excel (Table 1). The music was played using a Samsung Galaxy J5 and Shure 440 Hz headphones, with the volume standardized to the same level for all participants.
Table 1. List of selected songs and their order of play.

<table>
<thead>
<tr>
<th>Order</th>
<th>Title</th>
<th>Duration (min:s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fortnite OST–Battle Royale Menu Music (Rock Version) [31]a</td>
<td>3:50</td>
</tr>
<tr>
<td>2</td>
<td>Alone (Mashmallo) – Modified [32]b</td>
<td>3:19</td>
</tr>
<tr>
<td>3</td>
<td>Free Fire New EPIC Theme Song [33]c</td>
<td>3:56</td>
</tr>
<tr>
<td>4</td>
<td>Herobrine’s Life (Instrumental) [34]d</td>
<td>4:00</td>
</tr>
<tr>
<td>5</td>
<td>Olha a explosão (Mc Kevinho) – Modified [35]b</td>
<td>3:07</td>
</tr>
</tbody>
</table>

aThis song is part of the game Fortnite and was formerly played in the Battle Royale menu and when a player wins the Battle Royale mode. It was composed by Rom Di Prisco; all content belongs to Epic Games.
bThe original song was modified with Audacity (version 2.3.2; Audacity Team), an audio editor, to remove the voices.
cThis song is the theme song of Free Fire 2019.
dThis song is a Minecraft parody of the song “Something Just Like This” by The Chainsmokers and Coldplay.

Procedure

The attention task involved the child version of the ANT [36] under 2 conditions (with and without music). The child ANT was run using E-prime (version 2.0 professional; Psychology Software Tools) downloaded on a Samsung notebook from the webpage of Jin Fan [37]. All participants faced the laptop on the table in a comfortable, seated position (Figure 2). Prior to the ANT, the experimenter administered Conner’s Continuous Performance Test (CCPT) to the children. The entire procedure lasted approximately 1 hour and 30 minutes.

Figure 2. Experimental setup.

The first author (CGM) contacted parents or caregivers via telephone to arrange the experiment day. For the ADHD group, the experiment was scheduled on the same day as the participant’s psychiatrist appointment at the university hospital, or another agreed-upon day, to enhance adherence to the intervention. Children were individually escorted to a quiet office while their parents completed behavior scales and a sociodemographic questionnaire in the waiting room. It is crucial to highlight that this was a clinical sample. All children were under psychiatric monitoring and had a confirmed diagnosis of ADHD prior to participation.

For the control group, the experiment was scheduled on the most convenient day and location for caregivers, provided the child met all eligibility criteria. Screening for ADHD was conducted using the SNAP-IV and CBCL/6-18 scales, which were completed by the parents as a web form. When the experiment was held at a participant’s home, the child performed the task on a table in the quietest area of the house.

Additionally, all children and caregivers were asked to complete a semistructured questionnaire, which can be found in Multimedia Appendix 2, before starting the task.
Measures

Primary Outcome Measure

The ANT (child version) was designed to assess 3 attention networks (alerting, orienting, and conflict) within a single task framework based on the model described by Posner and Petersen [38]. In this version, participants are instructed to feed a central colorful fish by pressing a joystick button corresponding to the direction (left or right) in which it swims. The fish may appear alone or accompanied by other fish moving in the same or opposite direction (neutral, congruent, or incongruent stimuli) combined with various cuing conditions (no cue, central cue, double cue, and spatial cue) [38] (Figure 3).

Figure 3. Schematic of the child version of the ANT. In the actual task, the background color for every display is blue and the fish appear in yellow; the auditory feedback was used only in practice trials. RT: reaction time.

Originally, the task comprised 24 practice trials followed by 3 experimental blocks of 48 trials each. Since children completed the task twice (with and without music), practice rounds were administered separately. This procedure typically lasted approximately 45 minutes, including 5 minutes of practice and 15-minute rounds of 48 trials each, with 1-2 minute rest intervals. Psychometric properties of the ANT were assessed with a sample size of 40, yielding test-retest reliabilities of 0.52, 0.61, and 0.77 for the alerting, orienting, and conflict measures, respectively [39]. Additionally, with a sample size of 104, test-retest reliabilities of 0.36, 0.41, and 0.81 were reported for the alerting, orienting, and conflict measures, respectively [40].

Secondary Outcomes

The SNAP IV is a screening scale for ADHD and oppositional defiant disorder (ODD) based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria [29]. It consists of 26 items divided into subsets of symptoms (inattention, hyperactivity/impulsivity, and ODD) rated on a 4-point Likert scale ranging from 0 (not at all) to 3 (very much). Scores can be computed using 3 methods: averaging scores for each dimension, summing total scores, or counting the number of symptoms [29]. We used the symptom count to screen for ADHD, while the second and third methods were used for sample characterization. In a Brazilian sample, parental assessment of the SNAP IV demonstrated robust psychometric properties, with Cronbach α values of 0.94 and 0.92 for the inattention and hyperactivity scales, respectively [29].

The CBCL/6-18 is a self-report questionnaire assessing behaviors with 118 items scored as 0 (not true), 1 (somewhat or sometimes true), or 2 (very true or often true). Scores yield raw scores for 8 narrowband scales and 3 broadband scales, which are then transformed into t scores based on normative data [28]. The CBCL/6-18 aids in ruling out other pathologies potentially confounding ADHD diagnosis and establishing inclusion criteria for typically developing children. Internal consistencies as measured by Cronbach α for the problem scales range from 0.72 to 0.97 [28].

The CCPT is a computerized test measuring sustained attention and vigilance in individuals aged 6 years and older [41,42]. Performance metrics include measures of reaction times, errors, and response variability. Participants respond to letters displayed on a screen by pressing the spacebar, except when the letter X appears. The CCPT-2, chosen as a baseline attention measure, demonstrates good internal consistency (Cronbach α ranges from 0.64 to 0.96) and adequate test-retest reliability (coefficients range from 0.48 to 0.79) [42].
The Brazilian Economic Classification Criterion (CCEB) assesses the socioeconomic status of families based on household properties, educational attainment of the family head, and access to infrastructure [43]. Scores categorize families into socioeconomic levels from A to E. Considering that economically vulnerable children may exhibit more ADHD symptoms and externalizing disorders, this assessment ensures the appropriate pairing of groups to mitigate bias [44].

In addition to standardized instruments, a questionnaire gathered information on the children’s musical experience, preferences, and listening habits at home. Details can be found in Multimedia Appendix 3. At the study’s conclusion, the children were asked about their preferred testing condition.

**Statistical Analysis**

The sample size of 46 was determined using G*Power (version 3.0.10; Universität Düsseldorf), considering $\alpha = 0.05$, a medium effect size of 0.25, and a power of 90%. Statistical analyses were conducted using SPSS (version 22.0; IBM Corp) for personal computers. Descriptive statistics were used to calculate the mean and SD for sample characterization. Attention network scores (alerting, orienting, and conflict) for each participant were derived by subtracting error rates, including omissions, perseveration, and outliers, as outlined by Fan et al [40].

To compute orienting and alerting scores per participant, the mean reaction time (RT) per cue condition across flanker conditions was calculated (orienting = RT for spatial cue – RT for central cue; alerting = RT for double cue – RT for no cue condition). Conflict scores were obtained by computing the participant’s mean RT for each flanker condition across cue conditions (RT for incongruent – RT for congruent). Mean scores across participants were calculated for each network using an Excel macro obtained from Jin Fan [37]. Error rates were determined by averaging errors across all conditions (cue and flanker). Detailed calculations are illustrated in Figures 4 and 5, based on the Excel macro downloaded from Jin Fan [37].

**Figure 4.** Example of calculations of attentional networks for participant 1, session 2, using an Excel (Microsoft Corp) macro.
We used an ANOVA with a repeated measures design to test if the experimental conditions (music×no music), groups (control×ADHD), and their interactions were related to changes in ANT scores. We used 1 model for each attention network measure and error rate. To reduce potential biases arising from individual differences in attention on ANT scores, we included fine-grained age-corrected measures of attention errors from the CCPT. Errors of commission (ie, responding to a stimulus when one should not) and omission (not responding to a stimulus when one should) were entered as covariates in each model.

Ethical Approval

Ethical approval for this study was obtained from the University’s Ethics Review Committee (97425218.4.0000.5149). Written informed consent was provided by all parents or guardians, while minors provided written informed assent before participation in the trial. The study protocol was initially registered at ReBEC.gov (U1111-12589039). The study adhered to the Transparent Reporting of Evaluations with Non-randomized Designs (TREND) statement [45].

Results

Descriptive behavioral characteristics of the eligible participants are presented in Table 2. Additional information regarding the participants’ previous musical experiences and emotional connections with the music selections can be found in Multimedia Appendices 1-3. Nearly half of the children reported both familiarity with and enjoyment of the songs used in this study. Moreover, 57 of 76 children (75%) expressed a preference for taking the test while listening to music compared to the no-music condition.

Independent sample t tests (2-tailed) revealed no significant age differences between the control and ADHD groups ($t_{74}=0.47; P=.63$). Similarly, no significant disparities in socioeconomic status were observed between the groups ($t_{74}=-1.158; P=.25$).
Table 2. Demographic characteristics of participants (N=76).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>TD group (n=42), mean (SD)</th>
<th>ADHD group (n=34), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>11.0 (0.85)</td>
<td>10.9 (0.75)</td>
</tr>
<tr>
<td>Brazilian Criteria of Economic Classification(^a)</td>
<td>28.5 (11.32)</td>
<td>31.5 (11.15)</td>
</tr>
</tbody>
</table>

**SNAP-IV\(^b\) symptoms (\(t\) score\(^c\))**

- Inattention: 1 (1.97) vs 6 (2.35)
- Hyperactivity/impulsivity: 0.5 (0.84) vs 5 (1.85)
- CBCL-ADHD\(^d\) (\(t\) score\(^c\)): 44.7 (5.90) vs 61.6 (11.47)

**CCPT\(^e\) score**

- Omissions: 11.7 (9.5) vs 11.9 (8.6)
- Commissions: 25.0 (6.0) vs 25.4 (7.3)

\(^a\)For the Brazilian Criteria of Economic Classification, 0-16=class D and E, 17-22=class C2, 23-28=class C1, 29-37=class B2, 38-44=class B1, and 44-100=class A.

\(^b\)SNAP-IV: Swanson, Noland, and Pelham Scale IV. Each item is categorized as present (1 point, which means all answers are equivalent to 2, “quite a bit,” or 3, “very much”) or absent (0 points, which means all answers equivalent to 0, “not at all,” or 1, “just a little”). The cutoff point for screening ADHD is >5.

\(^c\)Mean \(t\) scores calculated with reference to Brazilian normative data [25].

\(^d\)CBCL-ADHD: Child Behavior Checklist for Attention Deficit/Hyperactivity Disorder.

\(^e\)CCPT: Conner’s Continuous Performance Test.

**Effect of Music on ANT and Error Rate**

Table 3 presents the means and SDs for each ANT measure across different conditions and groups. Repeated measures ANOVA models are shown in Table 4. We did not find significant main effects for music or the interaction between group and music for the alerting, orienting, and conflict attentional networks (all were nonsignificant, with \(P\) values ranging from .28 to .74). Regarding ANT errors, we found a significant main effect for music with a moderate effect size (\(F_{1,72}=9.83; P=.03; \eta^2_p=0.06\)) but not for the group\(\times\)music interaction (\(F_{1,72}=1.79; P=.18\)). Both the typically developing participants (mean 0.041, SD 0.036 vs mean 0.039, SD 0.049) and ADHD participants (mean 0.066, SD 0.058 vs mean 0.052, SD 0.042) made fewer errors in the ANT while listening to music.

Table 3. Mean reaction time and SDs for correct responses in each condition (music and no music) and in both groups (attention deficit/hyperactivity disorder and typically developing).

<table>
<thead>
<tr>
<th>Group and flanker condition</th>
<th>Reaction time with music (ms), mean (SD)</th>
<th>Reaction time without music (ms), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Central cue</td>
<td>Double cue</td>
</tr>
<tr>
<td></td>
<td>No cue</td>
<td>Orienting</td>
</tr>
<tr>
<td>Typically developing group</td>
<td>639 (120)</td>
<td>626 (266)</td>
</tr>
<tr>
<td>Congruent</td>
<td>692 (130)</td>
<td>619 (192)</td>
</tr>
<tr>
<td></td>
<td>614 (177)</td>
<td>635 (135)</td>
</tr>
<tr>
<td></td>
<td>675 (229)</td>
<td>612 (296)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>702 (106)</td>
<td>689 (343)</td>
</tr>
<tr>
<td></td>
<td>736 (201)</td>
<td>680 (62)</td>
</tr>
<tr>
<td></td>
<td>680 (177)</td>
<td>674 (75)</td>
</tr>
<tr>
<td></td>
<td>723 (109)</td>
<td>660 (203)</td>
</tr>
<tr>
<td>Neutral</td>
<td>621 (167)</td>
<td>597 (115)</td>
</tr>
<tr>
<td></td>
<td>664 (227)</td>
<td>590 (182)</td>
</tr>
<tr>
<td></td>
<td>634 (205)</td>
<td>612 (255)</td>
</tr>
<tr>
<td></td>
<td>657 (63)</td>
<td>598 (246)</td>
</tr>
<tr>
<td>Attention deficit/hyperactivity disorder group</td>
<td>686 (120)</td>
<td>668 (266)</td>
</tr>
<tr>
<td>Congruent</td>
<td>743 (130)</td>
<td>663 (192)</td>
</tr>
<tr>
<td></td>
<td>689 (120)</td>
<td>672 (266)</td>
</tr>
<tr>
<td></td>
<td>734 (130)</td>
<td>648 (192)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>763 (106)</td>
<td>722 (343)</td>
</tr>
<tr>
<td></td>
<td>796 (201)</td>
<td>693 (62)</td>
</tr>
<tr>
<td></td>
<td>759 (106)</td>
<td>708 (343)</td>
</tr>
<tr>
<td></td>
<td>781 (201)</td>
<td>710 (62)</td>
</tr>
<tr>
<td>Neutral</td>
<td>672 (167)</td>
<td>658 (115)</td>
</tr>
<tr>
<td></td>
<td>732 (227)</td>
<td>647 (182)</td>
</tr>
<tr>
<td></td>
<td>651 (167)</td>
<td>648 (115)</td>
</tr>
<tr>
<td></td>
<td>709 (227)</td>
<td>629 (182)</td>
</tr>
</tbody>
</table>
Motivation in achieving school tasks [51-53]. Children with ADHD, which corroborates studies on the role of motivation and self-regulation problems, which lead to the devaluing of rewards that are not immediate in comparison to listening to music, but this is only speculative. In the case of this study, we chose to use music that was familiar and preferred by most participants, so it is possible that the combination of these factors, music that induces emotions, plus the listener's current emotional state or desired emotional state, individuals tend to seek out and prefer music that aligns with their current emotional state or desired emotional state. The concept of affect-matching music refers to the idea that music on performance attention.

Also, since it was developed, the ANT has been widely used by the scientific community in diverse cultures and investigations (for, eg, anxiety, ADHD, bilingualism, borderline personality disorder, deafness, mindfulness training, schizophrenia, and time of day) [46,58], and the child variant [36] is the gold standard in this population, being engaging and visually stimulating. Thus, it was the best tool to assess the effect of music on performance attention.

Since it was developed, the ANT has been widely used by the scientific community in diverse cultures and investigations (for, eg, anxiety, ADHD, bilingualism, borderline personality disorder, deafness, mindfulness training, schizophrenia, and time of day) [46,58], and the child variant [36] is the gold standard in this population, being engaging and visually stimulating. Thus, it was the best tool to assess the effect of music on performance attention.

This study hypothesized that listening to music during testing may improve the attention performance of children with ADHD. However, our findings did not fully support this hypothesis. We found listening to music can improve performance accuracy by decreasing the number of errors, and this happened in both groups. The deficits in the attentional networks of children with ADHD assessed through the ANT are still controversial, and there are previous studies that also did not find differences in the efficiency of their networks when compared to children without ADHD [24,46]. Also, higher alertness seems to be associated with increased error rates [45,47,48], so music helped promote an optimal condition, that is, one that would not affect the accuracy of the attentional network, or it generated a weak effect size for the detection of differences.

The effects of music on cognitive performance are affected by motivation, especially if it is a favorite song [48]. In this study, most of the participants reported positive feelings (ie, liking the song and feelings of happiness) about the pieces of music used, and they preferred to perform the ANT listening to music. This may have increased the motivation to complete the task, consequently contributing to making fewer mistakes while listening to music, but this is only speculative. In the case of the children with ADHD, they may have lower levels of motivation and self-regulation problems, which lead to the devaluing of rewards that are not immediate in comparison to typically developing children [49,50]. Although we did not find a significant difference between the groups, our results suggest a tendency for the effect of music to be more significant in the group with ADHD, which corroborates studies on the role of motivation in achieving school tasks [51-53]. Children with ADHD, when motivated, are more likely to try harder when faced with difficulties or not to give up when something is difficult to finish or is not interesting to them [54]. So, it is important to understand that strategies that motivate these children can directly affect their performance of a task, and this does not necessarily have to be through an attentional route.

Also, previous studies demonstrated that when music has lyrics, it might impair performance attention [55,56], which could have been another important factor that contributed to the reduction of errors during the performance of the task in our study. Still, our musical stimulus was composed of songs that could have created an atmosphere to captivate the children, as 3 of the songs were from games that these children routinely played. This may have generated a feeling of reward and motivated the children more during the test.

The concept of affect-matching music refers to the idea that individuals tend to seek out and prefer music that aligns with their current emotional state or desired emotional state. Improvements in cognitive performance are facilitated by listening to affect-matching music [57]. On the other hand, music is also capable of inducing emotions [12,15]. In this study, we chose to use music that was familiar and preferred by most participants, so it is possible that the combination of these factors, music that induces emotions, plus the listener’s perception (emotion and arousal), led to our results.

Since it was developed, the ANT has been widely used by the scientific community in diverse cultures and investigations (for, eg, anxiety, ADHD, bilingualism, borderline personality disorder, deafness, mindfulness training, schizophrenia, and time of day) [46,58], and the child variant [36] is the gold standard in this population, being engaging and visually stimulating. Thus, it was the best tool to assess the effect of music on performance attention.

This study has some limitations. First, the results are only generalizable to the specific music used in this study and potentially to other music in the same genre and with a specific visual task in a laboratory setting. It will be necessary to carry out the same study with the same type of music while performing school tasks or in the classroom.

### Table 4. Score comparison for participants with typical development (TD; n=42) and attention deficit/hyperactivity disorder (ADHD; n=34) on the Attentional Network Test with and without music, controlling for Conner’s Continuous Performance Test omission and commission errors (repeated measures ANOVA).

<table>
<thead>
<tr>
<th>Measure</th>
<th>TD group score, mean (SD)</th>
<th>ADHD group, mean (SD)</th>
<th>Main effect of music</th>
<th>Interaction of music×group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No music</td>
<td>Music</td>
<td>No music</td>
<td>Music</td>
</tr>
<tr>
<td>Alerting</td>
<td>68.62 (52.67)</td>
<td>60.33 (81.28)</td>
<td>79.56 (54.64)</td>
<td>62.32 (64.75)</td>
</tr>
<tr>
<td>Orienting</td>
<td>26.69 (47.6)</td>
<td>24.29 (40.88)</td>
<td>38.62 (50.33)</td>
<td>44.03 (57.2)</td>
</tr>
<tr>
<td>Conflict</td>
<td>58.02 (47.06)</td>
<td>53.81 (38.28)</td>
<td>56.18 (47.35)</td>
<td>56.50 (38.28)</td>
</tr>
<tr>
<td>Error rate</td>
<td>0.041 (0.036)</td>
<td>0.039 (0.049)</td>
<td>0.066 (0.058)</td>
<td>0.052 (0.042)</td>
</tr>
</tbody>
</table>

*αp²=0.06.*

### Discussion

#### Principal Findings

The results across both ADHD and control groups revealed neither a significant main effect of music in attention networks, as indexed by the ANT, nor a significant interaction between music and group. However, a significant main effect was found in the overall number of errors during the ANT, suggesting listening to music decreases the error rate.

This study hypothesized that listening to music during testing may improve the attention performance of children with ADHD. However, our findings did not fully support this hypothesis. We found listening to music can improve performance accuracy by decreasing the number of errors, and this happened in both groups. The deficits in the attentional networks of children with ADHD assessed through the ANT are still controversial, and there are previous studies that also did not find differences in the efficiency of their networks when compared to children without ADHD [24,46]. Also, higher alertness seems to be associated with increased error rates [45,47,48], so music helped promote an optimal condition, that is, one that would not affect the accuracy of the attentional network, or it generated a weak effect size for the detection of differences.

The effects of music on cognitive performance are affected by motivation, especially if it is a favorite song [48]. In this study, most of the participants reported positive feelings (ie, liking the song and feelings of happiness) about the pieces of music used, and they preferred to perform the ANT listening to music. This may have increased the motivation to complete the task, consequently contributing to making fewer mistakes while listening to music, but this is only speculative. In the case of the children with ADHD, they may have lower levels of motivation and self-regulation problems, which lead to the devaluing of rewards that are not immediate in comparison to typically developing children [49,50]. Although we did not find a significant difference between the groups, our results suggest a tendency for the effect of music to be more significant in the group with ADHD, which corroborates studies on the role of motivation in achieving school tasks [51-53]. Children with ADHD, when motivated, are more likely to try harder when faced with difficulties or not to give up when something is difficult to finish or is not interesting to them [54]. So, it is important to understand that strategies that motivate these children can directly affect their performance of a task, and this does not necessarily have to be through an attentional route.

Also, previous studies demonstrated that when music has lyrics, it might impair performance attention [55,56], which could have been another important factor that contributed to the reduction of errors during the performance of the task in our study. Still, our musical stimulus was composed of songs that could have created an atmosphere to captivate the children, as 3 of the songs were from games that these children routinely played. This may have generated a feeling of reward and motivated the children more during the test.

The concept of affect-matching music refers to the idea that individuals tend to seek out and prefer music that aligns with their current emotional state or desired emotional state. Improvements in cognitive performance are facilitated by listening to affect-matching music [57]. On the other hand, music is also capable of inducing emotions [12,15]. In this study, we chose to use music that was familiar and preferred by most participants, so it is possible that the combination of these factors, music that induces emotions, plus the listener’s perception (emotion and arousal), led to our results.

Since it was developed, the ANT has been widely used by the scientific community in diverse cultures and investigations (for, eg, anxiety, ADHD, bilingualism, borderline personality disorder, deafness, mindfulness training, schizophrenia, and time of day) [46,58], and the child variant [36] is the gold standard in this population, being engaging and visually stimulating. Thus, it was the best tool to assess the effect of music on performance attention.

This study has some limitations. First, the results are only generalizable to the specific music used in this study and potentially to other music in the same genre and with a specific visual task in a laboratory setting. It will be necessary to carry out the same study with the same type of music while performing school tasks or in the classroom.

https://www.i-jmr.org/2024/1/e53869

(page number not for citation purposes)
Second, the sample of children with ADHD was recruited from only one clinical care setting, and this may have generated biases. Also, because it is a single clinical sample, the findings may not generalize to the broader population, limiting the external validity of the study.

**Conclusion**

Our findings, while preliminary, suggest that music does not appear to interfere with attentional networks. However, they do indicate that listening to music reduces the number of errors during directed attention tasks such as the ANT. Can similar results be observed during academic tasks? Could listening to music serve as a means to motivate children, thereby enhancing their engagement and accuracy in completing tasks? These questions warrant further investigation.

The motivational significance of a task plays a crucial role in channeling the additional effort required to sustain attention, potentially contributing to the reduction in errors observed. However, the effects of music on attention may vary among individuals with ADHD. While some children may find certain types of music beneficial for enhancing attention, others may find it disruptive. Therefore, it is essential to consider personal preferences and sensitivities when assessing the impact of music on attention in children with ADHD.

Ultimately, our research underscores the importance of exploring alternative and complementary treatments for ADHD that incorporate music, as it possesses intrinsic motivating potential and is readily accessible in people’s daily lives. Further studies are needed to deepen our understanding of how music can be effectively used to support attention and cognitive function in individuals with ADHD.

**Acknowledgments**

This study was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the National Institutes of Science and Technology in Responsible Neurotechnology (INCT Neurotec-R) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

**Data Availability**

The data analyzed during this study are available from the corresponding author on reasonable request.

**Authors' Contributions**

CGM designed and conducted the study and wrote the first version of the manuscript in consultation with DMM, JJP designed and conducted statistical analysis on the data, CGM structured and interpreted the results. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Assessment of emotional state.

[DOCX File, 113 KB - ijmr_v13i1e53869_app1.docx ]

**Multimedia Appendix 2**

Frequency (percentage) of answers to the questionnaire to assess the emotional relationship between listener and songs.

[DOCX File, 15 KB - ijmr_v13i1e53869_app2.docx ]

**Multimedia Appendix 3**

Frequency (percentage) of answers of preview musical experience questionnaire.

[DOCX File, 15 KB - ijmr_v13i1e53869_app3.docx ]

**References**


33. Free Fire New EPIC Theme Song. FreeFire Original. URL: https://www.youtube.com/watch?v=oCBMY0MsiWA [accessed 2024-06-21]


37. Cognitive neuroscience laboratory at QC. Queens College CUNY. URL: http://people.qc.cuny.edu/Faculty/Jin/Fan/Pages/Downloads.aspx [accessed 2024-03-23]


**Abbreviations**

ADHD: attention deficit/hyperactivity disorder  
ANT: Attention Network Test  
CBCL/6-18: Child Behavior Checklist for Ages 6-18  
CCCEB: Brazilian Economic Classification Criterion  
CCPT: Conner’s Continuous Performance Test  
K-SADS-PL: Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime  
ODD: oppositional defiant disorder  
RT: reaction time  
SNAP-IV: Swanson, Noland, and Pelham Scale IV  
TREND: Transparent Reporting of Evaluations with Non-randomized Designs

©Camila Guimarães Mendes, Jonas Jardim de Paula, Débora Marques Miranda. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.07.2024. 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.
Treating Spontaneous Pneumothorax Using an Innovative Surgical Technique Called Capnodissection Pleurectomy: Case Report

Ghaith Qsous, MRCS, MD; Prashanth Ramaraj, MSc, MRCS; Sanjeet Singh Avtaar Singh, MRCS, PhD; Philip Herd, MBChB; Nayandra Runveer Sooraj, MMed, FCA (SA); Malcolm Brodie Will, FRCS (CTh), PhD

1Department of Cardiothoracic Surgery, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom
2Imperial College School of Medicine, Imperial College London, London, United Kingdom
3Department of Anaesthesia, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Prashanth Ramaraj, MSc, MRCS
Department of Cardiothoracic Surgery
Royal Infirmary of Edinburgh
Little France Crescent
Old Dalkeith Road
Edinburgh, EH16 4SA
United Kingdom
Phone: 44 01315361000 ext 21076
Email: Prashanth.ramaraj@nhs.net

Abstract

Spontaneous pneumothorax is one of the most common conditions encountered in thoracic surgery. This condition can be treated conservatively or surgically based on indications and guidelines. Traditional surgical management includes pleurodesis (mechanical or chemical) in addition to bullectomy if the bullae can be identified. Mechanical pleurodesis is usually performed by surgical pleurectomy or pleural abrasion. In this case report, we present a case of a young patient with spontaneous pneumothorax who needed a surgical intervention. We performed a new, innovative surgical technique for surgical pleurectomy where we used carbon dioxide for dissection of the parietal pleura (capnodissection). This technique may provide similar efficiency to the traditional procedure but with less risk of bleeding and complications.


KEYWORDS

capnodissection; pleurectomy; VATS; video-assisted thorascopic surgery; novel technique; thoracic surgery; surgical innovation; pneumothorax; spontaneous pneumothorax; pleurodesis; management; bullectomy; bullae; young patient; lung diseases; chronic obstructive pulmonary disease; COPD; surgical treatment; male; capnothorax

Introduction

Spontaneous pneumothorax (SP) is a condition in which pneumothorax occurs without trauma or iatrogenic cause. It can be classified as a primary SP if there is no obvious underlying lung disease. The most common cause is usually a small bulla or bleb in the lung [1,2]. Comparatively, secondary SP happens due to underlying lung diseases such as chronic obstructive pulmonary disease [3]. The new British Thoracic Society (BTS) guidelines advise surgical treatment for SP at initial presentation if recurrence prevention is deemed important (eg, patients presenting with tension pneumothorax or those in high-risk occupations). Elective surgery should be considered for patients with a second ipsilateral or first contralateral pneumothorax [4].

The surgical treatment that is recommended by the BTS guidelines for SP is surgical pleurodesis with or without bullectomy [4]. There are 2 common ways to perform surgical pleurodesis: the first one is surgical pleurectomy and the second one is pleural abrasion. Surgical pleurectomy is considered more efficient, but it can be associated with an increased risk of bleeding and infection [5]. The novel surgical technique that we provide here can give a similar success rate but with less risk of complications such as bleeding or infection.

Case Presentation

Our patient was a young male individual who was previously healthy. He presented with a recurrence of an SP for the first time (2 SPs in total). The previous episode was treated conservatively 7 months prior, and his computed tomography
scan for this episode showed that he had small apical bullae. The decision was made to list the patient for elective surgical treatment, and after discussion with the patient, he was listed for a pleurectomy and bullectomy.

A standard anterior video-assisted thorascopic surgery (VATS) approach was taken. A small incision was made at the sixth intercostal space, and another small port site was created for the camera, which was later converted into the drain site (Figure 1). Carbon dioxide (CO$_2$) insufflation at 6-8 mm Hg on high flow was used to achieve capnothorax. A small anterior VATS incision was made at the sixth intercostal space, and the dissection of the parietal pleura was performed extrapleurally using Roberts forceps with a traditional technique. The forceps were exchanged for a curved metal sucker, and the CO$_2$ insufflation was attached at high flow and used to mobilize the whole parietal pleura, first from apex to inferior and then from posterior to anterior (Multimedia Appendix 1). The posterior parietal pleura was then excised off the ribs using thorascopic scissors 4 cm from the sympathetic chain posteriorly, 2 cm lateral to the internal mammary vein anteriorly, and 2 cm cranially to the diaphragm. Lastly, a bullectomy was performed using a manual stapler to excise the presumed culprit apical bullae seen on the computed tomography scan. Blood loss was minimal, approximately 50 mL, predominantly from VATS entry. Operative time was approximately 40 minutes. The postoperative care was routine, and the drain was removed after 48 hours. The patient was discharged on the third postoperative day.

Figure 1. Video-assisted thorascopic surgery (VATS) ports setup and surgical instruments.

**Ethical Considerations**

On the day of the operation, the patient completed a written consent form. The patient kindly agreed to the recording of the procedure and the utilization of his nonidentifiable data for this case report and publication, which was further discussed between the patient and GQ. Consent was sought by GQ and given by those in the operating theater for recording of the technique and publication as a case report.

**Discussion**

The use of CO$_2$ in thoracic surgery has increased significantly with the growing use of a minimally invasive approach. Capnothorax leads to better visualization by collapsing the lung and reduces the rate of complications [6,7]. In our department, we usually use CO$_2$ with robot-assisted thorascopic surgery and VATS for these reasons.

Surgical pleurodesis of SP is the recommended treatment in the BTS guidelines because it gives better long-term outcomes with less risk of recurrence in the future [4]. Surgical pleurectomy, in spite of it is efficiency, carries a risk of bleeding, infection, and reoperation [8]. Surgical pleural abrasion is another method that can be used for surgical pleurodesis. Chang et al [9] published the first systematic review and meta-analysis that compared surgical abrasion against surgical apical pleurectomy. They found that there is no difference in the recurrence, but pleural abrasion has a shorter length of stay in hospital, postoperative chest tube duration, and operative time and less surgical blood loss [9]. This may cause clinicians to consider a change of practice from surgical pleurectomy to abrasion. A systematic review of randomized controlled trials found that SPs managed with a chest drain alone had recurrence rates that ranged from 26.1% to 50.1%, whereas after VATS talc pleurodesis, these ranged from 0% to 3.2%. Alternative chemical pleurodesis can be achieved with tetracycline rather than talc, although recurrence rates were reported as ranging from 13% to 33.3% [10].

Our literature search did not find any studies in which capnodissection was used for pleurectomy as a treatment of SP. However, Dai, et al [11] recently published their findings for using CO$_2$ for visceral pleurectomy and decortication in patients with malignant mesothelioma. They found that the positive pressure of CO$_2$ can facilitate dissection of the visceral pleura, making the procedure easier while achieving an acceptable postoperative air leak and chest drain output. They concluded that capnodissection can be used in pleurectomy and decortication for patients with mesothelioma [11]. It should be...
taken into consideration that although the effect of capnodissection on gas exchange has not been rigorously studied, there is evidence to show that hypercarbia can result from CO₂ insufflation for capnothorax during VATS or robot-assisted thoracic surgery procedures [12]. This must be taken into account by surgeons and anesthetists when considering compensatory ventilator strategies, especially in patients with compromised gas exchange.

Our experience with the use of capnodissection for surgical pleurectomy was successful, and after 17 months from the procedure, the patient did not have any recurrence or complications. Moreover, this technique was not time-consuming (40-minute operative time), and the patient was discharged after 48 hours, with the surgeons noticing less pain in comparison to the traditional surgical pleurectomy, although pain is subjective. After VATS talc pleurodesis, the chest drain is typically removed no sooner than the second postoperative day, with discharge later that day. There is a theoretical risk of increased recurrence, as while the relatively atraumatic nature of this technique may reduce patient pain, it may also reduce the proinflammatory process required for pleurodesis and hence recurrence prevention [13]. More cases and longer follow-up are required to investigate the noninferiority of our technique to the traditional procedure.

Conclusions

In this case, capnodissection of the parietal pleura was a novel, safe, and successful technique that may decrease the risk of bleeding and postoperative pain.

Acknowledgments

The authors would like to thank the Imperial College London Library for access to article processing charges (APCs) waiver through the Imperial-JMIR unlimited Open Access (OA) deal.

Data Availability

Data regarding this paper have been deposited in the supplemental files. Corresponding author PR will enable further data sharing upon reasonable request.

Authors’ Contributions

GQ conceived the idea of the case report, assisted during this procedure, and lead the manuscript writing. PR was involved in manuscript writing, internal review, and internal editing. SSAS was involved in the internal review. PH and NRS were anesthetists for this case. MBW was the operating surgeon.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Video of capnodissection in action with surgeon narrative.
[MP4 File (MP4 Video), 17601 KB - ijmrv131e54497_app1.mp4]

References


Abbreviations

BTS: British Thoracic Society
CO2: carbon dioxide
SP: spontaneous pneumothorax
VATS: video-assisted thorascopic surgery

©Ghaith Qsous, Prashanth Ramaraj, Sanjeet Singh Avtaar Singh, Philip Herd, Nayandra Runveer Sooraj, Malcolm Brodie Will. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 21.06.2024. 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.
Case Report

Radiological Progression of Degenerative Cervical Myelopathy in a Clinically Stable Patient: Case Report

Rishi Umeria¹, MA, MB BChir; Oliver Mowforth¹, MA, MB BChir, MSt, MRCS; Munashe Veremu¹, BSc, MBBS; Benjamin Davies¹, MBChB, BSc, MRCS; Mark Kotter¹, MD, MPhil, PhD
Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Mark Kotter, MD, MPhil, PhD
Division of Neurosurgery
Department of Clinical Neurosciences
University of Cambridge
Addenbrooke’s Hospital
Hills Road
Cambridge, CB2 0QQ
United Kingdom
Phone: 44 1223 336946
Email: mrk25@cam.ac.uk

Abstract

Degenerative cervical myelopathy (DCM) is a common neurological condition, with disease progression that is both variable and difficult to predict. Here, we present a case of DCM in a gentleman in his late 60s with significant radiological disease progression without consequent change in clinical symptoms. The case serves as a reminder of an enduring medical aphorism that clinical history and examination should be prioritized above more complex data, such as imaging investigations. In addition, the case also highlights that guidelines should be contextualized within individual clinical circumstances.

KEYWORDS
degenerative cervical myelopathy; neurosurgery; radiology; magnetic resonance imaging

Introduction

Degenerative cervical myelopathy (DCM) is the most common cause of adult spinal cord dysfunction worldwide and is estimated to affect up to 2% of adults [1,2]. It arises secondary to degenerative pathology in the cervical spine, which leads to spinal cord compression. Cord compression may precipitate progressive neurological deficits including motor, sensory, and sphincter dysfunction [3].

The radiological progression of DCM does not always correlate well with its clinical progression [4-6]. Imaging findings such as the level and severity of spinal cord compression may not match the patient’s experience of their symptoms over time, thereby adding complexity to the clinician’s role of interpreting and reconciling clinical and radiological features of the disease [1,7].

Guidelines are clear in the event of symptom progression and advocate for surgery [8]. In contrast, the management of radiological progression without worsening symptoms is less defined and remains controversial. This report describes a case of the latter scenario, with the discussion focused on management.

Case Presentation

A retired man in his late 60s presented reporting a 2-year history of lower limb weakness, impaired balance, and calf aches. He associated this with a preceding episode of influenza. He had a background of long-standing lower back pain secondary to lumbar stenosis, for which he had previously undergone 3 surgical decompressions and atrial fibrillation, for which he took aspirin daily. His father had undergone a cervical laminectomy many years previously. The patient was otherwise fit and well, regularly cycled for exercise, and was a nonsmoker.

His symptoms resulted in a referral to neurology and investigations including magnetic resonance imaging (MRI) of his cervical spine. The MRI revealed multilevel degenerative changes from C3 to C7 in the cervical spine, on a background of a congenitally narrow spinal canal (Figure 1). He was
therefore referred to the neurosurgery department. Consent was obtained from the patient.

On assessment in the neurosurgery clinic, the patient was diagnosed with DCM. His modified Japanese Orthopaedic Association (mJOA) score was 14 (4+6+2+2); a score of 12 to 14 indicates moderate myelopathy. Further exploration of his symptoms revealed dysesthesia in the region of the left shoulder, hypoesthesia of the third and fourth digits of the left hand, difficulties with tandem walking, and bladder issues with minor episodes of urinary incontinence. He was reluctant to undergo surgery, and hence, it was agreed to monitor his symptoms with reassessment in the clinic, with a strong recommendation to consider surgery if there was further progression.

The appointment 5 months later demonstrated no further progression of his symptoms, with an unchanged mJOA score. A further 6 months later, at a third neurosurgical clinic appointment, the patient reported some deterioration in his condition; he had been finding cycling more difficult, felt lower back stiffness, had worsened pain, and reported a mild electric shock sensation on the right upper limb. His mJOA score remained unchanged. He underwent a repeat MRI of his cervical spine, with a view to considering surgical decompression. The MRI showed progression of the degenerative changes at C3 to C7 (Figure 2).

Nonetheless, at his subsequent follow-up appointment, the patient reported improvement in the symptoms and the mJOA score remained 14. He had, however, developed left L4 sciatica, which had become his main concern. Three months later at his fifth appointment, his mJOA score was 13 (4+5+2+2). He reported increasing work and concentration required to walk and climb stairs. Compared to the third appointment, there had been a slow, gradual clinical progression. At this stage, the patient was put on the waiting list for surgery, just before the national lockdown in the United Kingdom during the COVID-19 pandemic.

At 3 subsequent follow-up appointments over the succeeding 8 months, there was no further clinical progression of the patient’s DCM. Therefore, it was agreed to remove him from the waiting list for surgery and continue with an expectant approach. At the ninth appointment 6 months later, symptoms were again stable. At the 10th appointment, the patient reported some worsening back and lower limb pain, as well as some deterioration in his general mobility. A further MRI was therefore requested, which demonstrated significant radiological progression of his DCM compared to his MRI 5 years earlier (Figure 3).

At the 11th appointment, 5 months later, the symptoms had improved to the patient’s previous baseline, including balance, dexterity, numbness, and urinary urgency. His mJOA remained 13. The radiological progression was in stark contrast with the fact that the patient remained clinically stable. Therefore, it was agreed to continue with expectant management and continue close clinical follow-up.

Figure 1. T2-weighted sagittal magnetic resonance imaging demonstrating multilevel degenerative cervical spondylosis and disc degeneration.
Figure 2. T2-weighted sagittal MRI demonstrating the progression of multilevel degenerative changes in the cervical spine, including disc degeneration at C3-C7.

Figure 3. T2-weighted sagittal magnetic resonance imaging showing significant radiological worsening of cervical cord compression at C3-C7 levels despite unchanged symptoms and modified Japanese Orthopaedic Association clinical severity score.
Ethical Considerations

According to the National Health Service Health Research Authority, research ethics committee approval was not required for this work. Informed patient consent was obtained for publication.

Investigations

The first MRI showed multilevel degenerative changes of the neck on the background of a congenitally narrow spinal canal: C3/4—disc osteophyte complex causing central compression of the spinal cord; C4/5— moderate stenosis; C5/6—moderate stenosis; C6/7—severe compression, with associated T2 signal changes.

The second MRI showed disc and facet joint degenerative changes between C3 and C7; C3/4—central disc protrusion causing moderate spinal canal stenosis and compression of the spinal cord, which had progressed compared to the first MRI; C4/5—broad-based disc bulge causing mild spinal canal narrowing and indenting the undersurface of the cord; C5/6—broad-based disc bulge causing moderate spinal canal narrowing and mildly compressing the spinal cord, which had progressed compared to the first MRI; C6/7—broad-based disc bulge causing mild to moderate spinal canal narrowing, indenting the anterior surface of the spinal cord.

The third MRI scan showed multilevel degenerative changes from C3 to C7, with scoliotic deformity: C3/4—severe circumferential stenosis with cord compression; C4/5—significant circumferential stenosis with cord compression; C5/6—circumferential stenosis with cord compression; C6/7—circumferential stenosis with cord compression.

Differential Diagnosis

The differential diagnosis for a presentation of lower limb weakness, imbalance, and calf ache can be divided into upper and lower motor neuron patterns of weakness. Causes for the upper motor neuron pattern include pathologies in the brain and spinal cord. In the brain, this includes demyelinating disorders such as multiple sclerosis, vascular disorders such as stroke, space-occupying lesions such as a parasagittal meningioma or abscess, and motor neuron diseases such as amyotrophic lateral sclerosis. In the spinal cord, this includes demyelinating disorders such as transverse myelitis, myelopathies such as DCM, space-occupying lesions, for example, tumor or abscess, trauma, syringomyelia, and spinal stenosis (spinal claudication).

Causes for the lower motor neuron pattern include drugs such as alcohol, metabolic disorders such as vitamin B₁₂ deficiency, diabetes mellitus, inherited disorders such as Charcot-Marie-Tooth, infections such as HIV or syphilis, and autoimmune disorders such as vasculitis and chronic inflammatory demyelinating polyneuropathy. The causes of calf aches include trauma, vascular disorders such as peripheral vascular disease (intermittent claudication), and inflammatory disorders such as myositis.

Treatment

Since being diagnosed 5 years ago, an expectant approach has been taken in the management of this patient. International DCM management guidelines recommend surgical management for moderate, severe, or progressive DCM; however, for mild DCM, the optimal treatment strategy remains undefined, with a recommendation of either surgery or supervised nonsurgical management [8]. In this patient’s case, there is moderate DCM, with an mJOA score of 13, with significant radiological progression but clinical and symptomatic stability across serial assessments.

Outcome and Follow-Up

The patient continues close neurosurgical follow-up, currently at 12 monthly intervals, alongside careful safety netting advice.

Discussion

Principal Findings

Degenerative changes in the cervical spine include disc herniation, osteophytosis, ligament hypertrophy, and ossification [9]. DCM is a clinical syndrome that arises when these changes result in spinal cord compression that is associated with symptoms, which may include neck pain or stiffness, limb pain or weakness, urinary incontinence, decreased manual dexterity, imbalance, or falls [1]. In this patient’s case, there was also a risk factor for congenital stenosis of the cervical canal.

A challenging aspect of the management of DCM is how to deal with symptoms changing over time and correlating this with evolving imaging findings. Another challenging aspect is the important decision on the timing of any surgical management.

The guidelines advise clinicians to take a structured, consistent approach to management. Each person with DCM requires consideration of individual factors, which may mean that clinical judgment or patient preferences result in deviation from guidelines in some circumstances. The patient was diagnosed with moderate DCM; the mJOA score for the patient was initially 14 and then decreased and remained stable at 13.

Strict application of the guidelines would lead to a recommendation for surgical intervention. However, while the mJOA includes consideration of upper and lower limb motor function, upper limb sensory function, and sphincter function, it does not capture all symptoms and clinical features. For example, in this case, limb pain was not captured. Nonetheless, it is a validated scoring tool for the assessment of functional status and is responsive to changes in the severity of DCM [10].

The complexity of this case requires a nuanced approach to management. Surgical intervention within DCM is primarily aimed at halting symptom progression; however, without symptom worsening, the decision of when to operate becomes more complicated. This patient’s symptoms were managed expectantly with nonsurgical interventions such as physical therapy as tolerated, oral analgesics, and neuropathic agents for any acute pain flares. Urinary symptoms were stable and were not actively managed. Waiting to operate at an older age may increase the risk of further complications, and this possibility...
should be explained to the patient. This should be part of a shared decision-making approach, where patients are empowered to make decisions through collaboration with their clinicians with the understanding that, in the context of a chronic disease like DCM, this decision will likely be revisited [11].

The factors driving the disconnect between clinical and radiological severity are unknown. Nonetheless, a model proposed by Davies et al [12] postulates that DCM is a function of (1) mechanical stress, (2) duration of injury, and (3) individual vulnerability. Using this model, a scenario of limited clinical progression and significant radiological progression over time could be explained by decreased vulnerability to injury.

An individual’s vulnerability to DCM comprises primary protective mechanisms such as genetics and age, in addition to adaptive protective mechanisms, such as autoregulation of spinal cord perfusion, functional reserve capacity, and nutritional status [12]. For example, certain genotypes are associated with increased regenerative capacity, such as the HIF-1A polymorphism rs11549467 [13]. This polymorphism is associated with susceptibility to DCM and its clinical features, including severity measured by mJOA.

Furthermore, adaptive protective mechanisms such as autoregulation of spinal cord perfusion may minimize ischemic injury. Decreased blood flow can result in blood-spinal cord barrier dysfunction, leading to microglia activation, neuroinflammation, and neuronal apoptosis [14]. In addition to the ischemia precipitating apoptosis [14], dysregulation of the autoregulatory system can occur from mechanical cord compression in DCM [15-17]. It is possible that dysregulation occurs to a lesser degree in some individuals, such as this patient. Furthermore, reserve capacity within the central nervous system [17-19] refers to resilience in the neurological system to account for the disconnect between the clinical phenotype and underlying histological pathology. In the context of DCM, cervical spinal cord compression and injury are initially asymptomatic, and the radiological changes affecting the spinal cord represent, at best, a proxy for the clinical presentation of the condition [12].

Conclusions
In summary, individuals may have different vulnerabilities and protective mechanisms that may account for the disconnect between clinical and radiological features of DCM. The message is therefore to treat the patient rather than treating the findings from the imaging. This is especially important, given that asymptomatic cervical spinal cord compression is common in the general population [2,7].

Acknowledgments
We gratefully acknowledge support from the Cambridge NIHR Brain Injury MedTech Cooperative. MK is funded by a National Institute for Health Research (NIHR) Clinician Scientist Award (CS-2015-15-023). BD is supported by an NIHR Clinical Doctoral Research Fellowship. OM is supported by an NIHR Academic Clinical Fellowship. This report is independent research arising from a Clinician Scientist Award (CS-2015-15-023) supported by the NIHR. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health. No generative AI was used in the generation of this manuscript.

Data Availability
All data generated or analyzed during this study are included in this published article.

Authors’ Contributions
RU, OM, MV, BD, and MK contributed to the writing and review of this article.

Conflicts of Interest
Author MK is a trustee of Myelopathy.org, without any financial interests.

References


Abbreviations

DCM: degenerative cervical myelopathy
mJOA: modified Japanese Orthopaedic Association
MRI: magnetic resonance imaging
Visual “Scrollytelling”: Mapping Aquatic Selfie-Related Incidents in Australia

Introduction

Selfies are a modern, yet preventable cause of injury and death [1]. Medical responses and retrieval of persons, often in challenging terrain, burdens emergency medicine practitioners. To help prevent this issue, this study aimed to visualize selfie-related incidents globally by initially creating a scrollytelling map story overlayed on a satellite map of the incidents in Australia. This type of visual storytelling technique using a world map helps illustrate the spatial context of this public health issue.

Methods

Overview

Incident data were acquired via publicly accessible news reports and a Wikipedia repository [2] and cleaned and prepared in Excel (Microsoft Corp). Incidents in aquatic areas (eg, coastal locations and inland waterfalls) were included; those in other settings (eg, falls from artificial structures and incidents involving trains) were excluded. Entries for each incident were created using associated media reports, incident types bring falls or drowns. Map coordinates were obtained by locating the incident using Google Maps and inputted into a coordinate finder using the Mapbox Location Helper [3]. Mapbox Studio [4] was used to create a custom map. A satellite template was chosen to best display the geographic context surrounding each selfie incident. The data set was imported into the Mapbox Studio custom map, which populated the data layer onto the map. A heat map setting was chosen for the data. To create the “scrollytelling” map story, the Mapbox storytelling template available from GitHub was used [5]. The primary input is a story comprising sections (chapters), each associated with a particular view of a map, enabling the user to “scroll” down the web page, and the resulting output is a zoomed-in view of a specific case layered on the map. The data and corresponding map visualization are hosted on GitHub and are published with GitHub pages.

Ethical Considerations

Ethics approval was not required due to the use of publicly available media reports and a Wikipedia repository to create the map.

Results

The publicly accessible map and data can be viewed on the web [6]. The heat map displays a worldwide overview of 104 cases from June 2014 through August 2023 (Figure 1 [7-9]). Once the user scrolls, the map displays 11 chapters, each showing an individual selfie-related incident in Australia including 9 deaths and 2 serious injuries. All cases reportedly involved emergency services, 3 of which occurred in the same location: Diamond Bay, Vaucluse, a suburb of Sydney, New South Wales. Two cases occurred at Gibraltar Falls, Australian Capital Territory. Further cases in Australia that are included in the map story are detailed in Multimedia Appendix 1.
Discussion

Our heat map of media-based incident data provides a globally applicable visual representation of the selfie-incident phenomenon in Australia. Using a scrolling story template, overlayed on top of a map, selfie incidents can be illustrated in a geographic context. It is clear from the heat map that certain locations worldwide (nations such as India) and in Australia (cities such as Sydney) require specific and targeted prevention strategies to attenuate the incidence of selfie incidents, which is in line with their specific topographical realities.

Future research in this space should seek to further ascertain the burden on emergency and retrieval services by evaluating response times, resource allocation, retrieval or rescue methods, and health care costs associated with treating selfie-related injuries. Geographic disparities in service usage and response times, terrain, and retrieval or rescue methods should be identified and added to the visual map.

Understanding the geographic distribution and burden of selfie-related incidents is essential for designing targeted public awareness campaigns, improving safety regulations, and optimizing the allocation of resources for emergency and retrieval services.

The main limitation results from the use of media cases as the basis of these incident reports. It is not possible to acquire precise latitude and longitude coordinates. The coordinates provide the best approximation using details provided in media reports and analyzing the geography described in the report. Nevertheless, this map provides a good overview of the geographic nature of selfie-related incidents. Given the ethical limitations regarding the use of coronial data and identification of individual incidents, publicly available information of this nature remains the most appropriate data source.

In conclusion, selfie-related incidents present a significant geographic challenge for emergency services and retrievalists due to the inherent geographic context surrounding this type of event. Mapping selfie events may be a useful method of analyzing and tracking these phenomena and may be of benefit to emergency managers and land managers in collaborating to attenuate this issue.
Acknowledgments
This work was funded by a University of New South Wales (UNSW) university postgraduate award provided to SC. AP is funded by a National Health and Medical Research Council (NHMRC) Emerging Leadership Fellowship (grant APP2009306).

Data Availability
All data are publicly available and can be found in a GitHub repository [10].

Conflicts of Interest
None declared.

Multimedia Appendix 1
Table displaying cases used in the scrollytelling map.

References
5. mapbox / storytelling. GitHub. URL: https://github.com/mapbox/storytelling [accessed 2023-09-06]