# Interactive Journal of Medical Research

Impact Factor (2022): 2.0
Volume 13 (2024) ISSN 1929-073X

## Contents

### Original Papers

## Viewpoints

## Reviews

## Original



## Short Paper



## Corrigenda and Addendas

## Case Reports



## Research Letter



## Commentary

Original Paper

# Predicting the Effectiveness of a Mindfulness Virtual Community Intervention for University Students: Machine Learning Model

Christo El Morr[1], PhD; Farideh Tavangar[2], MSc; Farah Ahmad[1], PhD; Paul Ritvo[3], PhD; MVC Team[4]

[1]School of Health Policy and Management, York University, Toronto, ON, Canada

[2]Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[3]Kinesiology & Health Science, York University, Toronto, ON, Canada

[4]See Acknowledgments

**Corresponding Author:**
Christo El Morr, PhD
School of Health Policy and Management
York University
4700 Keele Street
Toronto, ON, M3J 1P3
Canada
Phone: 1 426 736 2100 ext 22053
Email: elmorr@yorku.ca

## 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

***(Interact J Med Res 2024;13:e50982)*** 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

## Introduction

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 communities, 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

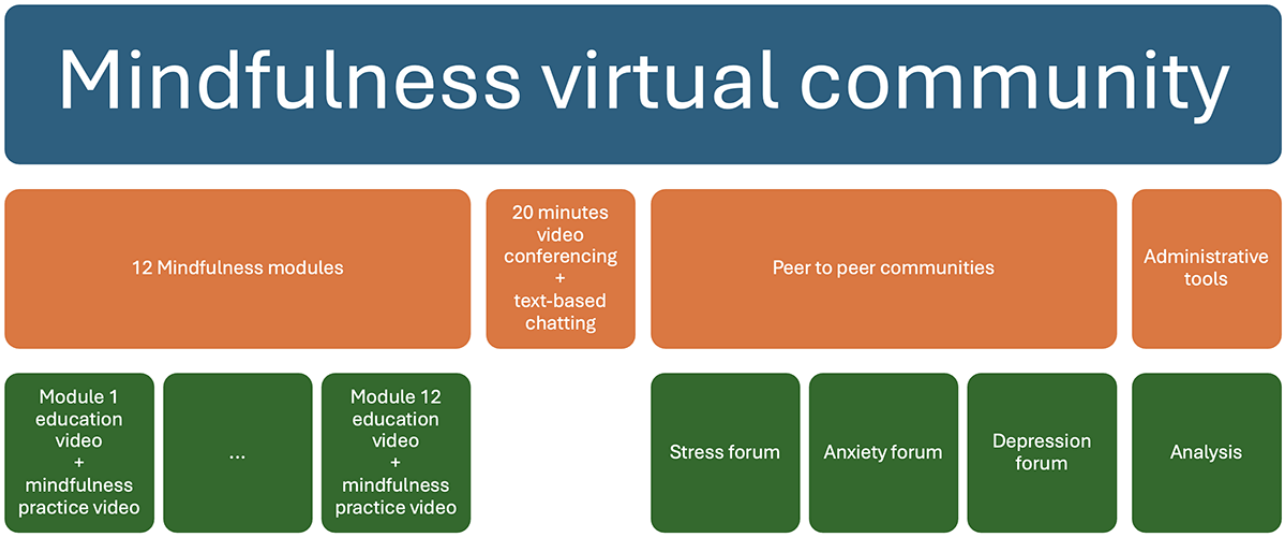
This is a retrospective analysis, where we analyzed an anonymized data set. The data were deidentified, and consent was obtained during the RCT; no further consent was sought for this secondary data analysis since nonidentifiable data were used. The data set was collected via an RCT described in detail 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.

**Figure 1.** The mindfulness virtual community design.




## 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 1.** Algorithms and their corresponding optimal hyperparameters found by grid search (data set without videos).

| Algorithm | Parameters |
|---|---|
| **Logistic regression** | |
| PHQ-9[a] | C=1, penalty=l1, solver=liblinear |
| BAI[b] | C=1, penalty=l1, solver=liblinear |
| PSS[c] | C=1, penalty=l1, solver=liblinear |
| **Support vector machine** | |
| PHQ-9 | C=10, γ=0.1, kernel=rbf |
| BAI | C=10, γ=0.1, kernel=rbf |
| PSS | C=10, γ=0.1, kernel=rbf |
| **Random forest** | |
| PHQ-9 | Max_features=auto, n_estimators=500, max_depth=8, criterion=entropy |
| BAI | Max_features=auto, n_estimators=500, max_depth=8, criterion=gini |
| PSS | Max_features=auto, n_estimators=500, max_depth=8, criterion=gini |
| **Decision tree** | |
| PHQ-9 | Max_leaf_nodes=59, random_state=42, min_samples_split=2, criterion=entropy |
| BAI | Max_leaf_nodes=56, random_state=42, min_samples_split=2, criterion=entropy |
| PSS | Max_leaf_nodes=16, random_state=42, min_samples_split=2, criterion=entropy |
| **K-nearest neighbor** | |
| PHQ-9 | N_neighbors=2, weight=distance, leaf size=27, P=1 |
| BAI | N_neighbors=2, weight=distance, leaf size=1, P=1 |
| PSS | N_neighbors=1, weight=dniform, leaf size=1, P=1 |
| **Adaptive boosting** | |
| PHQ-9 | n-estimators=5000, max_depth=3, learning rate=0.5 |
| BAI | n-estimators=5000, max_depth=3, learning rate=0.9 |
| PSS | n-estimators=500, max_depth=3, learning rate=0.9 |
| **Gradient boosting** | |
| PHQ-9 | Learning rate=0.05, max depth=6, n-estimators=100, subsample=0.9, max_features=none, min_samples_split=2 |
| BAI | Learning rate=0.02, max depth=10, n-estimators=1000, subsample=1.0, max_features=none, min_samples_split=2 |
| PSS | Learning rate=0.01, max depth=6, n-estimators=1000, subsample=0.9, max_features=sqrt, min_samples_split=2 |

[a]PHQ-9: Patient Health Questionnaire-9.

[b]BAI: Beck Anxiety Inventory.

[c]PSS: Perceived Stress Scale.

**Table 2.** Algorithms and their corresponding optimal hyperparameters found by grid search (data set with exposure to videos).

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

[a]PHQ-9: Patient Health Questionnaire-9.

[b]BAI: Beck Anxiety Inventory.

[c]PSS: 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.

**Table 3.** Characteristics of participants at baseline (N=209).

| Characteristics | Values |
|---|---|
| **Sex, n (%)** | |
| Male | 56 (26.8) |
| Female | 153 (73.2) |
| **Marital status, n (%)** | |
| Married | 17 (8.1) |
| Other | 192 (91.9) |
| **Ethnicity, n (%)** | |
| White | 44 (21.1) |
| Non-White | 165 (78.9) |
| **Language, n (%)** | |
| English | 136 (65.1) |
| Other | 73 (34.9) |
| **Country of birth, n (%)** | |
| Canada | 119 (56.9) |
| Other | 90 (43.1) |
| **Education, n (%)** | |
| High school diploma or General Education Development or college degree or certificate program | 182 (87.1) |
| Bachelor degree | 27 (12.9) |
| **Self-reported general health, n (%)** | |
| Poor or fair | 43 (20.6) |
| Good or very good or excellent | 166 (79.4) |
| Age (years), median (IQR) | 21 (19-23) |
| Average number of hours at work per week, median (IQR) | 10 (0-18) |
| Total number of mindfulness videos watched, median (IQR) | 16 (9-30) |

## 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 AdaBoost (AUC=0.92 and accuracy=0.92) and gradient boosting (AUC=0.87 and accuracy=0.87), which outperformed the remaining classifiers. Two classifiers, gradient boosting and DT, gained the perfect accuracy and AUC (AUC=1 and accuracy=1) for predicting a clinically significant reduction in PSS, followed by the near-perfect scores for SVM and AdaBoost (AUC=0.99 and accuracy=0.99). 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=0.91 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.

| Algorithm | AUC[a] | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| **Logistic regression** | | | | |
| PHQ-9[b] | 0.64 | 0.66 | 0.57 | 0.72 |
| BAI[c] | 0.75 | 0.75 | 0.75 | 0.75 |
| PSS[d] | 0.73 | 0.73 | 0.73 | 0.74 |
| **Support vector machine** | | | | |
| PHQ-9 | 0.81 | 0.80 | 0.90 | 0.75 |
| BAI | 0.77 | 0.77 | 0.79 | 0.75 |
| PSS | 0.96 | 0.96 | 1.0 | 0.91 |
| **Random forest** | | | | |
| PHQ-9 | 0.78 | 0.76 | 0.87 | 0.69 |
| BAI | 0.93 | 0.93 | 0.86 | 1.0 |
| PSS | 0.99 | 0.99 | 1.0 | 0.97 |
| **Decision tree** | | | | |
| PHQ-9 | 0.84 | 0.81 | 0.96 | 0.72 |
| BAI | 0.84 | 0.83 | 0.93 | 0.75 |
| PSS | 1.0 | 1.0 | 1.0 | 1.0 |
| **K-nearest neighbor** | | | | |
| PHQ-9 | 0.82 | 0.80 | 0.91 | 0.72 |
| BAI | 0.78 | 0.78 | 0.68 | 0.88 |
| PSS | 0.96 | 0.96 | 1.0 | 0.91 |
| **Adaptive boosting** | | | | |
| PHQ-9 | 0.82 | 0.80 | 0.91 | 0.72 |
| BAI | 0.92 | 0.92 | 0.89 | 0.94 |
| PSS | 0.99 | 0.99 | 1.0 | 0.97 |
| **Gradient boosting** | | | | |
| PHQ-9 | 0.85 | 0.83 | 0.91 | 0.78 |
| BAI | 0.87 | 0.87 | 0.86 | 0.88 |
| PSS | 1.0 | 1.0 | 1.0 | 1.0 |

[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 5.** Classification report of the machine learning algorithms for outcomes (data set with exposure to videos).

| Algorithm | AUC[a] | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| **Logistic regression** | | | | |
| PHQ-9[b] | 0.62 | 0.63 | 0.57 | 0.67 |
| BAI[c] | 0.60 | 0.60 | 0.61 | 0.59 |
| PSS[d] | 0.79 | 0.80 | 0.85 | 0.74 |
| **Support vector machine** | | | | |
| PHQ-9 | 0.78 | 0.75 | 0.96 | 0.61 |
| BAI | 0.93 | 0.93 | 0.86 | 1.00 |
| PSS | 1.00 | 1.00 | 1.00 | 1.00 |
| **Random forest** | | | | |
| PHQ-9 | 0.84 | 0.83 | 0.87 | 0.81 |
| BAI | 0.90 | 0.90 | 0.86 | 0.94 |
| PSS | 1.00 | 1.00 | 1.00 | 1.00 |
| **Decision tree** | | | | |
| PHQ-9 | 0.84 | 0.81 | 0.96 | 0.72 |
| BAI | 0.80 | 0.80 | 0.86 | 0.75 |
| PSS | 0.97 | 0.97 | 1.00 | 0.94 |
| **K-nearest neighbor** | | | | |
| PHQ-9 | 0.81 | 0.78 | 0.96 | 0.67 |
| BAI | 0.84 | 0.83 | 0.89 | 0.78 |
| PSS | 0.99 | 0.99 | 1.00 | 0.97 |
| **Adaptive boosting** | | | | |
| PHQ-9 | 0.84 | 0.81 | 0.96 | 0.72 |
| BAI | 0.93 | 0.93 | 0.89 | 0.97 |
| PSS | 1.00 | 1.00 | 1.00 | 1.00 |
| **Gradient boosting** | | | | |
| PHQ-9 | 0.89 | 0.88 | 0.96 | 0.83 |
| BAI | 0.91 | 0.92 | 0.86 | 0.97 |
| PSS | 1.00 | 1.00 | 1.00 | 1.00 |

[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, FRCPC; Louise Hartley, PhD; Amin Mawani, PhD; Kwame McKenzie, MD, FRCPC; 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

1. American College Health Association-National College Health Assessment II: Canadian reference group executive summary spring 2013. American College Health Association. 2013. URL: https://tinyurl.com/vxn2swse [accessed 2024-04-18]
2. American College Health Association-National College Health Assessment II: Canadian reference group executive summary spring 2016. American College Health Association. 2016. URL: https://tinyurl.com/bdfpv8b4 [accessed 2024-04-18]
3. Pirbaglou M, Cribbie R, Irvine J, Radhu N, Vora K, Ritvo P. Perfectionism, anxiety, and depressive distress: evidence for the mediating role of negative automatic thoughts and anxiety sensitivity. J Am Coll Health 2013;61(8):477-483 [FREE Full text] [doi: 10.1080/07448481.2013.833932] [Medline: 24152025]
4. Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]
5. Han B, Compton WM, Eisenberg D, Milazzo-Sayre L, McKeon R, Hughes A. Prevalence and mental health treatment of suicidal ideation and behavior among college students aged 18-25 years and their non-college-attending peers in the United States. J Clin Psychiatry 2016;77(6):815-824 [FREE Full text] [doi: 10.4088/JCP.15m09929] [Medline: 27232194]
6. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, et al. WHO World Mental Health Surveys International College Student project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]
7. The human face of mental health and mental illness in Canada 2006. Government of Canada. 2006. URL: https://www.phac-aspc.gc.ca/publicat/human-humain06/pdf/human_face_e.pdf [accessed 2024-04-18]
8. Allen R, Kannangara C, Vyas M, Carson J. European university students' mental health during Covid-19: exploring attitudes towards Covid-19 and governmental response. Curr Psychol 2022:1-14 [FREE Full text] [doi: 10.1007/s12144-022-02854-0] [Medline: 35194357]
9. Koelen JA, Mansueto AC, Finnemann A, de Koning L, van der Heijde CM, Vonk P, et al. COVID-19 and mental health among at-risk university students: a prospective study into risk and protective factors. Int J Methods Psychiatr Res 2022;31(1):e1901 [FREE Full text] [doi: 10.1002/mpr.1901] [Medline: 34932250]
10. Zapata-Ospina JP, Patiño-Lugo DF, Vélez CM, Campos-Ortiz S, Madrid-Martínez P, Pemberthy-Quintero S, et al. Mental health interventions for college and university students during the COVID-19 pandemic: a critical synthesis of the literature. Rev Colomb Psiquiatr (Engl Ed) 2021;50(3):199-213 [FREE Full text] [doi: 10.1016/j.rcpeng.2021.04.001] [Medline: 34479842]
11. Ungar T. The health care payment game is rigged. National Post. 2015 Apr 28. URL: https://nationalpost.com/opinion/thomas-ungar-the-health-care-payment-game-is-rigged [accessed 2024-05-03]
12. Lim KL, Jacobs P, Ohinmaa A, Schopflocher D, Dewa CS. A new population-based measure of the economic burden of mental illness in Canada. Chronic Dis Can 2008;28(3):92-98 [FREE Full text] [Medline: 18341763]
13. Bayram N, Bilgel N. The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. Soc Psychiatry Psychiatr Epidemiol 2008;43(8):667-672 [FREE Full text] [doi: 10.1007/s00127-008-0345-x] [Medline: 18398558]

14. Pfeffer A. Ontario campus counsellors say they're drowning in mental health needs. CBC. 2019. URL: https://www.cbc.ca/news/canada/ottawa/mental-health-ontario-campus-crisis-1.3771682 [accessed 2019-07-07]
15. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
16. Keng SL, Smoski MJ, Robins CJ. Effects of mindfulness on psychological health: a review of empirical studies. Clin Psychol Rev 2011;31(6):1041-1056 [FREE Full text] [doi: 10.1016/j.cpr.2011.04.006] [Medline: 21802619]
17. Grossman P, Niemann L, Schmidt S, Walach H. Mindfulness-based stress reduction and health benefits. A meta-analysis. J Psychosom Res 2004;57(1):35-43 [FREE Full text] [doi: 10.1016/S0022-3999(03)00573-7] [Medline: 15256293]
18. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009;15(5):593-600 [FREE Full text] [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
19. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
20. Vøllestad J, Nielsen MB, Nielsen GH. Mindfulness- and acceptance-based interventions for anxiety disorders: a systematic review and meta-analysis. Br J Clin Psychol 2012;51(3):239-260 [FREE Full text] [doi: 10.1111/j.2044-8260.2011.02024.x] [Medline: 22803933]
21. Eberth J, Sedlmeier P. The effects of mindfulness meditation: a meta-analysis. Mindfulness 2012;3(3):174-189 [FREE Full text] [doi: 10.1007/s12671-012-0101-x]
22. Sedlmeier P, Eberth J, Schwarz M, Zimmermann D, Haarig F, Jaeger S, et al. The psychological effects of meditation: a meta-analysis. Psychol Bull 2012;138(6):1139-1171. [doi: 10.1037/a0028168] [Medline: 22582738]
23. Bender JL, Jimenez-Marroquin MC, Ferris LE, Katz J, Jadad AR. Online communities for breast cancer survivors: a review and analysis of their characteristics and levels of use. Support Care Cancer 2013;21(5):1253-1263 [FREE Full text] [doi: 10.1007/s00520-012-1655-9] [Medline: 23179491]
24. El Morr C. Mobile virtual communities in healthcare: managed self care on the move. 2007 Presented at: Telehealth '07: The Third IASTED International Conference on Telehealth; May 31, 2007-June 1, 2007; Montreal, Canada. [doi: 10.4018/978-1-59904-002-8.ch105]
25. Morr CE, Kawash J. Mobile virtual communities research: a synthesis of current trends and a look at future perspectives. Int J Web Based Communities 2007;3(4):386-403 [FREE Full text] [doi: 10.1504/ijwbc.2007.015865]
26. Ahmad F, El Morr C, Ritvo P, Othman N, Moineddin R, MVC Team. An eight-week, web-based mindfulness virtual community intervention for students' mental health: randomized controlled trial. JMIR Ment Health 2020;7(2):e15520 [FREE Full text] [doi: 10.2196/15520] [Medline: 32074061]
27. Ritvo P, Ahmad F, El Morr C, Pirbaglou M, Moineddin R, MVC Team. A mindfulness-based intervention for student depression, anxiety, and stress: randomized controlled trial. JMIR Ment Health 2021;8(1):e23491 [FREE Full text] [doi: 10.2196/23491] [Medline: 33264098]
28. El Morr C, Ritvo P, Ahmad F, Moineddin R, MVC Team. Effectiveness of an 8-week web-based mindfulness virtual community intervention for university students on symptoms of stress, anxiety, and depression: randomized controlled trial. JMIR Ment Health 2020;7(7):e18595 [FREE Full text] [doi: 10.2196/18595] [Medline: 32554380]
29. El Morr C. Virtual communities, machine learning and IoT: opportunities and challenges in mental health research. Int J Extreme Autom Connectivity Healthc 2019;1(1):4-11 [FREE Full text] [doi: 10.4018/ijeach.2019010102]
30. Carter S, Greenberg J, Funes CJ, Macklin EA, Vranceanu AM. Effects of a mind-body program on symptoms of depression and perceived stress among adults with neurofibromatosis type 2 who are deaf: a live-video randomized controlled trial. Complement Ther Med 2021;56:102581 [FREE Full text] [doi: 10.1016/j.ctim.2020.102581] [Medline: 33197677]
31. Eskildsen A, Dalgaard VL, Nielsen KJ, Andersen JH, Zachariae R, Olsen LR, et al. Cross-cultural adaptation and validation of the Danish consensus version of the 10-item Perceived Stress Scale. Scand J Work Environ Health 2015;41(5):486-490 [FREE Full text] [doi: 10.5271/sjweh.3510] [Medline: 26111225]
32. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient Health Questionnaire. JAMA 1999;282(18):1737-1744 [FREE Full text] [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
33. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56(6):893-897. [doi: 10.1037//0022-006x.56.6.893] [Medline: 3204199]
34. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [Medline: 6668417]
35. Hornstein S, Forman-Hoffman V, Nazander A, Ranta K, Hilbert K. Predicting therapy outcome in a digital mental health intervention for depression and anxiety: a machine learning approach. Digit Health 2021;7:20552076211060659 [FREE Full text] [doi: 10.1177/20552076211060659] [Medline: 34868624]
36. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med Care 2004;42(12):1194-1201 [FREE Full text] [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]

37. Goldsworthy S, Palmer S, Latour JM, McNair H, Cramp M. A systematic review of effectiveness of interventions applicable to radiotherapy that are administered to improve patient comfort, increase patient compliance, and reduce patient distress or anxiety. Radiography (Lond) 2020;26(4):314-324 [FREE Full text] [doi: 10.1016/j.radi.2020.03.002] [Medline: 32245711]
38. Leentjens AFG, Dujardin K, Marsh L, Richard IH, Starkstein SE, Martinez-Martin P. Anxiety rating scales in Parkinson's disease: a validation study of the Hamilton anxiety rating scale, the Beck anxiety inventory, and the hospital anxiety and depression scale. Mov Disord 2011;26(3):407-415 [FREE Full text] [doi: 10.1002/mds.23184] [Medline: 21384425]
39. Wang C, Zhao H, Zhang H. Chinese college students have higher anxiety in new semester of online learning during COVID-19: a machine learning approach. Front Psychol 2020;11:587413 [FREE Full text] [doi: 10.3389/fpsyg.2020.587413] [Medline: 33343461]
40. Na KS, Cho SE, Geem ZW, Kim YK. Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. Neurosci Lett 2020;721:134804 [FREE Full text] [doi: 10.1016/j.neulet.2020.134804] [Medline: 32014516]
41. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
42. Moroz N, Moroz I, D'Angelo MS. Mental health services in Canada: barriers and cost-effective solutions to increase access. Healthc Manage Forum 2020;33(6):282-287 [FREE Full text] [doi: 10.1177/0840470420933911] [Medline: 32613867]
43. Roberts R, Golding J, Towell T, Reid S, Woodford S, Vetere A, et al. Mental and physical health in students: the role of economic circumstances. Br J Health Psychol 2010;5(3):289-297 [FREE Full text] [doi: 10.1348/135910700168928]
44. Trends in private and public funding in Canadian colleges, 2019/2020. Statistics Canada. 2023. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/220120/dq220120c-eng.htm [accessed 2023-03-26]
45. Pisaniello MS, Asahina AT, Bacchi S, Wagner M, Perry SW, Wong ML, et al. Effect of medical student debt on mental health, academic performance and specialty choice: a systematic review. BMJ Open 2019;9(7):e029980 [FREE Full text] [doi: 10.1136/bmjopen-2019-029980] [Medline: 31270123]
46. Boyles JD, Ahmed B. Does student debt affect dental students' and dentists' stress levels? Br Dent J 2017;223(8):601-606 [FREE Full text] [doi: 10.1038/sj.bdj.2017.876] [Medline: 29074930]
47. Gurevich E, El Hassan B, El Morr C. Equity within AI systems: what can health leaders expect? Healthc Manage Forum 2023;36(2):119-124 [FREE Full text] [doi: 10.1177/08404704221125368] [Medline: 36226507]
48. Kundi B, El Morr C, Gorman R, Dua E. Artificial intelligence and bias: a scoping review. In: Yampolskiy RV, editor. AI and Society: Tensions and Opportunities. New York: CRC Press Taylor & Francis; 2023:199-213.


## 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


*Edited by T de Azevedo Cardoso; submitted 18.07.23; peer-reviewed by M Musker, D Carvalho; comments to author 24.11.23; revised version received 05.12.23; accepted 05.04.24; published 13.05.24.*

*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*
*Interact J Med Res 2024;13:e50982*
*URL: https://www.i-jmr.org/2024/1/e50982*
*doi:10.2196/50982*
*PMID:38578872*

Viewpoint

# Designing mHealth Apps to Incorporate Evidence-Based Techniques for Prolonging User Engagement

Rebecca Monachelli[1], BS; Sharon Watkins Davis[1], MPA; Allison Barnard[1], BA; Michelle Longmire[1], MD; John P Docherty[2], MD; Ingrid Oakley-Girvan[1,3], MPH, PhD

[1]Medable Inc, Palo Alto, CA, United States

[2]Weill Cornell Medical College, White Plains, NY, United States

[3]The 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.

***(Interact J Med Res 2024;13:e51974)*** doi:10.2196/51974

**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.

| Task or questionnaire | Frequency | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 |
|---|---|---|---|---|---|---|---|
| Sign up | Once | Task 1 | N/A[a] | N/A | N/A | N/A | N/A |
| Demographics (age, gender, education, housing, race, ethnicity, location, health condition, and notification) | Once | Task 2 | N/A | N/A | N/A | N/A | N/A |
| Technology competence questions (TAM3[b] CANX,[c] worries about pressing the wrong button and device preference) | Twice | Task 3[d] | Task 2[e] | N/A | N/A | N/A | N/A |
| Technology competence video combined with practice questions (video on how to answer questions and practice questions) | Once | N/A | Task 1[d,e] | N/A | N/A | N/A | N/A |
| Health literacy questions (BRIEF[f], own health knowledge, clinical trial knowledge, and BMI) | Twice | N/A | N/A | Task 1[d,ed] | Task 2[e] | N/A | N/A |
| 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) | Once | N/A | N/A | N/A | Task 1[d,e] | N/A | N/A |
| Clinical trials question (temperature scale regarding participant's feelings about study participation, Figure S1 in Multimedia Appendix 1) | Twice | N/A | N/A | N/A | N/A | Task 1[d,e] | Task 2 |
| Clinical trials video (National Institute of Diabetes and Digestive and Kidney Diseases video: "Why Should I Join a Clinical Trial?") | Once | N/A | N/A | N/A | N/A | N/A | Task 1[d,e] |
| Study complete | Once | N/A | N/A | N/A | N/A | N/A | Task 3 |

[a]N/A: not applicable.

[b]TAM3: Technology Acceptance Model 3.

[c]CANX: Computer Anxiety.

[d]Set up with an automated morning reminder: "You have new tasks available today in the Patient Engagement app!"

[e]Set-up with an automated 8 pm reminder: "You have uncompleted tasks in the Patient Engagement app. Please finish them before midnight!"

[f]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 health 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 - ijmr_v13i1e51974_app1.docx ]

## References


1. Newzoo global mobile market report 2021. Newzoo. 2021. URL: https://newzoo.com/insights/trend-reports/newzoo-global-mobile-market-report-2021-free-version [accessed 2023-07-18]
2. Purnell JQ, Griffith J, Eddens KS, Kreuter MW. Mobile technology, cancer prevention, and health status among diverse, low-income adults. Am J Health Promot 2014;28(6):397-402 [FREE Full text] [doi: 10.4278/ajhp.120816-ARB-396] [Medline: 24200336]
3. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
4. Dorsey ER, Kluger B, Lipset CH. The new normal in clinical trials: decentralized studies. Ann Neurol 2020;88(5):863-866. [doi: 10.1002/ana.25892] [Medline: 32869367]
5. Sarraju A, Seninger C, Parameswaran V, Petlura C, Bazouzi T, Josan K, et al. Pandemic-proof recruitment and engagement in a fully decentralized trial in atrial fibrillation patients (DeTAP). NPJ Digit Med 2022;5(1):80 [FREE Full text] [doi: 10.1038/s41746-022-00622-9] [Medline: 35764796]
6. Sessa C, Cortes J, Conte P, Cardoso F, Choueiri T, Dummer R, et al. The impact of COVID-19 on cancer care and oncology clinical research: an experts' perspective. ESMO Open 2022;7(1):100339 [FREE Full text] [doi: 10.1016/j.esmoop.2021.100339] [Medline: 34953404]
7. Patel S, Goldsack JC, Cordovano G, Downing A, Fields KK, Geoghegan C, et al. Advancing digital health innovation in oncology: priorities for high-value digital transformation in cancer care. J Med Internet Res 2023;25:e43404 [FREE Full text] [doi: 10.2196/43404] [Medline: 36598811]
8. Davis SW, Oakley-Girvan I. mHealth education applications along the cancer continuum. J Cancer Educ 2015;30(2):388-394. [doi: 10.1007/s13187-014-0761-4] [Medline: 25482319]
9. DiMasi JA, Smith Z, Oakley-Girvan I, Mackinnon A, Costello M, Tenaerts P, et al. Assessing the financial value of decentralized clinical trials. Ther Innov Regul Sci 2023;57(2):209-219 [FREE Full text] [doi: 10.1007/s43441-022-00454-5] [Medline: 36104654]
10. Bruce CR, Harrison P, Nisar T, Giammattei C, Tan NM, Bliven C, et al. Assessing the impact of patient-facing mobile health technology on patient outcomes: retrospective observational cohort study. JMIR Mhealth Uhealth 2020;8(6):e19333 [FREE Full text] [doi: 10.2196/19333] [Medline: 32589161]
11. Tuvesson H, Eriksén S, Fagerström C. mHealth and engagement concerning persons with chronic somatic health conditions: integrative literature review. JMIR Mhealth Uhealth 2020;8(7):e14315 [FREE Full text] [doi: 10.2196/14315] [Medline: 32706686]
12. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
13. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
14. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022;9(1):e30766 [FREE Full text] [doi: 10.2196/30766] [Medline: 35099398]
15. Serrano KJ, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
16. Trifan A, Oliveira M, Oliveira JL. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. JMIR Mhealth Uhealth 2019;7(8):e12649 [FREE Full text] [doi: 10.2196/12649] [Medline: 31444874]
17. Kelders SM, van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
18. Madujibeya I, Lennie T, Aroh A, Chung ML, Moser D. Measures of engagement with mHealth interventions in patients with heart failure: scoping review. JMIR Mhealth Uhealth 2022;10(8):e35657 [FREE Full text] [doi: 10.2196/35657] [Medline: 35994345]
19. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]

20. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
21. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022;10(2):221 [FREE Full text] [doi: 10.3390/healthcare10020221] [Medline: 35206837]
22. Oakley-Girvan I, Lavista JM, Miller Y, Davis S, Acle C, Hancock J, et al. Evaluation of a mobile device survey system for behavioral risk factors (SHAPE): app development and usability study. JMIR Form Res 2019;3(1):e10246 [FREE Full text] [doi: 10.2196/10246] [Medline: 30684441]
23. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and e-health: a scoping review. Telemed J E Health 2022;28(6):768-780 [FREE Full text] [doi: 10.1089/tmj.2021.0176] [Medline: 34637651]
24. Oakley-Girvan I, Docherty JP. A new approach to enhancing engagement in eHealth apps. Interact J Med Res 2022;11(2):e38886 [FREE Full text] [doi: 10.2196/38886] [Medline: 36279587]
25. Lepore SJ, Rincon MA, Buzaglo JS, Golant M, Lieberman MA, Bass SB, et al. Digital literacy linked to engagement and psychological benefits among breast cancer survivors in internet-based peer support groups. Eur J Cancer Care (Engl) 2019;28(4):e13134 [FREE Full text] [doi: 10.1111/ecc.13134] [Medline: 31318132]
26. Haun J, Noland-Dodd V, Varnes J, Graham-Pole J, Rienzo B, Donaldson P. Testing the BRIEF health literacy screening tool. Fed Pract 2009;26(12):24-31 [FREE Full text]
27. Wellness 101 – how to improve your overall health. St. Luke's University Health Network YouTube page. 2019. URL: https://www.youtube.com/watch?v=_nuDp-fded8 [accessed 2024-03-01]
28. Adult BMI calculator. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/english_bmi_calculator/bmi_calculator.html [accessed 2023-07-18]
29. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008;39(2):273-315 [FREE Full text] [doi: 10.1111/j.1540-5915.2008.00192.x]
30. Rodgers GP. Why should I join a clinical trial? National Institute of Diabetes and Digestive and Kidney Disease YouTube page. 2019. URL: https://www.youtube.com/watch?v=36Sd8WpgR94 [accessed 2024-03-01]
31. 5 ways to make the most of your doctor visit. National Institute on Aging YouTube page. URL: https://www.youtube.com/watch?v=BincCVl-YsI [accessed 2023-07-18]
32. Are clinical trials safe? National Cancer Institute. URL: https://www.cancer.gov/about-cancer/treatment/clinical-trials/patient-safety [accessed 2023-07-19]
33. What are the risks and benefits of participating in clinical research? National Institute of Mental Health. URL: https://www.nimh.nih.gov/news/media/2021/what-are-the-risks-and-benefits-of-participating-in-clinical-research [accessed 2023-07-19]
34. 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 [accessed 2023-07-19]
35. Clinical research: benefits, risks, and safety. National Institute on Aging. URL: https://www.nia.nih.gov/health/clinical-research-benefits-risks-and-safety [accessed 2023-07-19]
36. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
37. Sim I. Mobile devices and health. N Engl J Med 2019;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
38. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
39. Moon Z, Zuchowski M, Moss-Morris R, Hunter MS, Norton S, Hughes LD. Disparities in access to mobile devices and e-health literacy among breast cancer survivors. Support Care Cancer 2022;30(1):117-126 [FREE Full text] [doi: 10.1007/s00520-021-06407-2] [Medline: 34236506]
40. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]


## Abbreviations

**mHealth:** mobile health
**PCN:** Patient Care Network

*Edited by G Eysenbach, T de Azevedo Cardoso; submitted 18.08.23; peer-reviewed by S Amagai, B Oh, C Latkin; comments to author 17.10.23; revised version received 14.11.23; accepted 27.02.24; published 26.03.24.*

*Please cite as:*
*Monachelli R, Davis SW, Barnard A, Longmire M, Docherty JP, Oakley-Girvan I*
*Designing mHealth Apps to Incorporate Evidence-Based Techniques for Prolonging User Engagement*
*Interact J Med Res 2024;13:e51974*
*URL: https://www.i-jmr.org/2024/1/e51974*
*doi:10.2196/51974*
*PMID:38416858*

Viewpoint

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

Catherine L Saunders[1], MBBS, PhD

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.

***(Interact J Med Res 2024;13:e53311)*** doi:10.2196/53311

**KEYWORDS**

lesbian; gay; bisexual; trans; LGBTQ+; routine data; England; United Kingdom; health; viewpoint; sexual orientation; health services; infrastructure data; policy; gender; health outcome; epidemiology; risk prediction; risk

## 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 using routine data from the United Kingdom [54]. In our recent work collaboratively exploring LGBTQ+ research priorities, intersectionality (understanding the interdependent and overlapping systems of discrimination and disadvantage) was identified as an area of research need; and race, ethnicity, and socioeconomic inequalities were particularly highlighted [55]. Larger sample sizes mean that intersectionality can now begin to be explored quantitatively [56]; newer longitudinal collections are providing additional insight [57].

Again, it is not just biomedical models that are important. Routine educational data sets have been important in highlighting the higher levels of bullying experienced by young LGBTQ+ people in schools [21].

However, we also know that on its own research describing inequalities experienced by LGBTQ+ adults will not lead directly to improved outcomes. Process measures of care quality

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

1. National LGBT survey: research report. Government Equalities Office. 2018. URL: https://www.gov.uk/government/publications/national-lgbt-survey-summary-report [accessed 2024-04-10]
2. King M, Semlyen J, Tai SS, Killaspy H, Osborn D, Popelyuk D, et al. A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. BMC Psychiatry 2008;8:70 [FREE Full text] [doi: 10.1186/1471-244X-8-70] [Medline: 18706118]
3. McDermott E, Nelson R, Weeks H. The politics of LGBT+ health inequality: conclusions from a UK scoping review. Int J Environ Res Public Health 2021;18(2):826 [FREE Full text] [doi: 10.3390/ijerph18020826] [Medline: 33478019]
4. Mendenhall E, Newfield T, Tsai AC. Syndemic theory, methods, and data. Soc Sci Med 2022;295:114656 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114656] [Medline: 34949486]
5. Tsai AC. Syndemics: a theory in search of data or data in search of a theory? Soc Sci Med 2018;206:117-122 [FREE Full text] [doi: 10.1016/j.socscimed.2018.03.040] [Medline: 29628175]
6. Meyer IH. Minority stress and mental health in gay men. J Health Soc Behav 1995;36(1):38-56. [Medline: 7738327]
7. Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol 2006;6:35 [FREE Full text] [doi: 10.1186/1471-2288-6-35] [Medline: 16872487]
8. Elliott MN, Kanouse DE, Burkhart Q, Abel GA, Lyratzopoulos G, Beckett MK, et al. Sexual minorities in England have poorer health and worse health care experiences: a national survey. J Gen Intern Med 2015;30(1):9-16 [FREE Full text] [doi: 10.1007/s11606-014-2905-y] [Medline: 25190140]
9. Semlyen J, Hagger-Johnson G. Sampling frame for sexual minorities in public health research. J Public Health (Oxf) 2017;39(3):644 [FREE Full text] [doi: 10.1093/pubmed/fdw078] [Medline: 27613765]
10. McManus S. Sexual Orientation Research Phase 1: A Review of Methodological Approaches. Edinburgh: Scottish Executive Social Research; 2003.
11. Davies M, Moon G, Sabel CE. Navigating the LGB data landscape: a review of appropriate secondary data sources for sexuality and substance use research in the UK. Int J Environ Res Public Health 2022;19(3):1329 [FREE Full text] [doi: 10.3390/ijerph19031329] [Medline: 35162352]

12. Equality Act 2010. legislation.gov.uk. London: HMSO; 2010. URL: https://www.legislation.gov.uk/ukpga/2010/15/contents [accessed 2024-04-17]
13. Question development. Office for National Statistics. 2021. URL: https://www.ons.gov.uk/census/planningforcensus2021/questiondevelopment [accessed 2024-04-10]
14. Ford J, Sowden S, Olivera J, Bambra C, Gimson A, Aldridge R, et al. Transforming health systems to reduce health inequalities. Future Healthc J 2021;8(2):e204-e209 [FREE Full text] [doi: 10.7861/fhj.2021-0018] [Medline: 34286186]
15. Knight HE, Deeny SR, Dreyer K, Engmann J, Mackintosh M, Raza S, et al. Challenging racism in the use of health data. Lancet Digit Health 2021;3(3):e144-e146 [FREE Full text] [doi: 10.1016/S2589-7500(21)00019-4] [Medline: 33549513]
16. Robbins HA, Alcala K, Swerdlow AJ, Schoemaker MJ, Wareham N, Travis RC, et al. Comparative performance of lung cancer risk models to define lung screening eligibility in the United Kingdom. Br J Cancer 2021;124(12):2026-2034 [FREE Full text] [doi: 10.1038/s41416-021-01278-0] [Medline: 33846525]
17. Khor S, Haupt EC, Hahn EE, Lyons LJL, Shankaran V, Bansal A. Racial and ethnic bias in risk prediction models for colorectal cancer recurrence when race and ethnicity are omitted as predictors. JAMA Netw Open 2023;6(6):e2318495 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.18495] [Medline: 37318804]
18. Berner AM, Lund J, Saunders CL. Tackling the complexity of gender bias in primary care. Br J Gen Pract 2021;71(708):296-297 [FREE Full text] [doi: 10.3399/bjgp21X716177] [Medline: 34319880]
19. University of Essex, Institute for Social and Economic Research. Understanding society: 9th release. UK Data Service. 2023. URL: http://doi.org/10.5255/UKDA-Series-2000053 [accessed 2024-04-10]
20. University College London, UCL Institute of Education, Centre for Longitudinal Studies. Millennium cohort study: 15th release. UK Data Service. 2023. URL: http://doi.org/10.5255/UKDA-Series-2000031 [accessed 2024-04-10]
21. Henderson M. Understanding bullying experiences among sexual minority youths in England. Centre for Longitudinal Studies Working paper/8. 2015. URL: https://cls.ucl.ac.uk/wp-content/uploads/2017/04/CLS-WP-20158.pdf [accessed 2024-04-10]
22. Health survey England additional analyses - health and health-related behaviours of lesbian, gay and bisexual adults. NHS Digital. 2021. URL: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-england-additional-analyses/lesbian-gay-and-bisexual-adults [accessed 2024-04-10]
23. Russell M, NatCen Social Research, Senior M. National survey of sexual attitudes and lifestyles, 2010-2012: secure access. UK Data Service. 2021. URL: https://datacompass.lshtm.ac.uk/id/eprint/633/ [accessed 2024-04-10]
24. Saunders CL, Berner A, Lund J, Mason AM, Oakes-Monger T, Roberts M, et al. Demographic characteristics, long-term health conditions and healthcare experiences of 6333 trans and non-binary adults in England: nationally representative evidence from the 2021 GP patient survey. BMJ Open 2023;13(2):e068099 [FREE Full text] [doi: 10.1136/bmjopen-2022-068099] [Medline: 36731935]
25. Saunders CL, MacCarthy S, Meads C, Massou E, Mant J, Saunders AM, et al. Long-term conditions among sexual minority adults in England: evidence from a cross-sectional analysis of responses to the English GP patient survey. BJGP Open 2021;5(5):BJGPO.2021.0067 [FREE Full text] [doi: 10.3399/BJGPO.2021.0067] [Medline: 34465579]
26. Saunders CL, Meads C, Abel GA, Lyratzopoulos G. Associations between sexual orientation and overall and site-specific diagnosis of cancer: evidence from two national patient surveys in England. J Clin Oncol 2017;35(32):3654-3661 [FREE Full text] [doi: 10.1200/JCO.2017.72.5465] [Medline: 28945501]
27. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. Nature 2018;562(7726):203-209 [FREE Full text] [doi: 10.1038/s41586-018-0579-z] [Medline: 30305743]
28. Our Future Health. 2023. URL: https://ourfuturehealth.org.uk/about-us/ [accessed 2024-04-10]
29. Brown J, Pfeiffer RM, Shrewsbury D, O'Callaghan S, Berner AM, Gadalla SM, et al. Prevalence of cancer risk factors among transgender and gender diverse individuals: a cross-sectional analysis using UK primary care data. Br J Gen Pract 2023;73(732):e486-e492 [FREE Full text] [doi: 10.3399/BJGP.2023.0023] [Medline: 37365010]
30. Jackson SS, Brown J, Pfeiffer RM, Shrewsbury D, O'Callaghan S, Berner AM, et al. Analysis of mortality among transgender and gender diverse adults in England. JAMA Netw Open 2023;6(1):e2253687 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.53687] [Medline: 36716027]
31. Rimes KA, Ion D, Wingrove J, Carter B. Sexual orientation differences in psychological treatment outcomes for depression and anxiety: national cohort study. J Consult Clin Psychol 2019;87(7):577-589. [doi: 10.1037/ccp0000416] [Medline: 31219292]
32. Henson KE, Elliss-Brookes L, Coupland VH, Payne E, Vernon S, Rous B, et al. Data resource profile: national cancer registration dataset in England. Int J Epidemiol 2020;49(1):16-16h [FREE Full text] [doi: 10.1093/ije/dyz076] [Medline: 31120104]
33. Equality and human rights monitor 2023. Equality and Human Rights Commission. 2023. URL: https://www.gov.uk/government/publications/equality-and-human-rights-monitor-2023 [accessed 2024-04-10]
34. Saunders CL, Massou E. Using UK Biobank for sexual minority health research. J Comp Eff Res 2021;10(13):1001-1010 [FREE Full text] [doi: 10.2217/cer-2021-0052] [Medline: 34231371]

35. Booker CL, Rieger G, Unger JB. Sexual orientation health inequality: evidence from understanding society, the UK longitudinal household study. Prev Med 2017;101:126-132 [FREE Full text] [doi: 10.1016/j.ypmed.2017.06.010] [Medline: 28601620]
36. Semlyen J. Recording sexual orientation in the UK: pooling data for statistical power. Am J Public Health 2017;107(8):1215-1217 [FREE Full text] [doi: 10.2105/AJPH.2017.303910] [Medline: 28657782]
37. Jamieson A, Cross H, Arthur S, Nambiar K, Llewellyn CD. Patient sexual orientation and gender identity disclosure. Clin Teach 2020;17(6):669-673 [FREE Full text] [doi: 10.1111/tct.13182] [Medline: 33217175]
38. Brooks H, Llewellyn CD, Nadarzynski T, Pelloso FC, Guilherme FDS, Pollard A, et al. Sexual orientation disclosure in health care: a systematic review. Br J Gen Pract 2018;68(668):e187-e196 [FREE Full text] [doi: 10.3399/bjgp18X694841] [Medline: 29378698]
39. Pollard A, Bradley J, Cooper M, Sadhwani S, Northam A, Llewellyn C. The NHS England fundamental information standard for monitoring sexual orientation. Br J Gen Pract 2019;69(679):94-95 [FREE Full text] [doi: 10.3399/bjgp19X701213] [Medline: 30705020]
40. Cass review- quantitative research. Cass Review. URL: https://cass.independent-review.uk/cass-review-quantitive-research/ [accessed 2024-04-10]
41. Mansfield KL, Puntis S, Soneson E, Cipriani A, Geulayov G, Fazel M. Study protocol: the OxWell school survey investigating social, emotional and behavioural factors associated with mental health and well-being. BMJ Open 2021;11(12):e052717 [FREE Full text] [doi: 10.1136/bmjopen-2021-052717] [Medline: 34880020]
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]
43. Hu Y, Denier N. Sexual orientation identity mobility in the United Kingdom: a research note. Demography 2023;60(3):659-673 [FREE Full text] [doi: 10.1215/00703370-10769825] [Medline: 37257022]
44. MacCarthy S, Saunders CL, Elliott MN. Increased reporting of sexual minority orientation from 2009 to 2017 in England and implications for measuring sexual minority health disparities. LGBT Health 2020;7(7):393-400 [FREE Full text] [doi: 10.1089/lgbt.2019.0181] [Medline: 32758045]
45. Cross H, Llewellyn CD. A decline in patient disclosure of heterosexuality in the English general practice patient survey: a longitudinal analysis of cross-sectional data. Fam Pract 2020;37(5):661-667 [FREE Full text] [doi: 10.1093/fampra/cmaa033] [Medline: 32270180]
46. Yarrow E, Duschinsky R, Saunders CL. The gender variance scale: developing and piloting a new tool for measuring gender diversity in survey research. J Gend Stud 2023;33(2):244-256 [FREE Full text] [doi: 10.1080/09589236.2023.2277463]
47. Saunders CL, Lund J, Mason AM, Roberts M, Smith J, Duschinsky R. The health and healthcare outcomes of trans and/or non-binary adults in England: protocol for an analysis of responses to the 2021 GP patient survey. Sexes 2022;3(3):325-335 [FREE Full text] [doi: 10.3390/sexes3030025]
48. Elliott MN, Dahlhamer JM, MacCarthy S, Beckett MK, Orr N, Guerino P, et al. Using ancillary sociodemographic data to identify sexual minority adults among those responding "something else" or "don't know" to sexual orientation questions. Med Care 2019;57(12):e87-e95 [FREE Full text] [doi: 10.1097/MLR.0000000000001190] [Medline: 31415342]
49. Hagger-Johnson G, Taibjee R, Semlyen J, Fitchie I, Fish J, Meads C, et al. Sexual orientation identity in relation to smoking history and alcohol use at age 18/19: cross-sectional associations from the Longitudinal Study of Young People in England (LSYPE). BMJ Open 2013;3(8):e002810 [FREE Full text] [doi: 10.1136/bmjopen-2013-002810] [Medline: 23985386]
50. Shahab L, Brown J, Hagger-Johnson G, Michie S, Semlyen J, West R, et al. Sexual orientation identity and tobacco and hazardous alcohol use: findings from a cross-sectional English population survey. BMJ Open 2017;7(10):e015058 [FREE Full text] [doi: 10.1136/bmjopen-2016-015058] [Medline: 29074508]
51. Meads C, Zeeman L, Sherriff N, Aranda K. Prevalence of alcohol use amongst sexual and gender minority (LGBTQ+) communities in the UK: a systematic scoping review. Alcohol Alcohol 2023;58(4):346-356 [FREE Full text] [doi: 10.1093/alcalc/agad029] [Medline: 37114766]
52. Boehmer U, Miao X, Ozonoff A. Cancer survivorship and sexual orientation. Cancer 2011;117(16):3796-3804 [FREE Full text] [doi: 10.1002/cncr.25950] [Medline: 21557209]
53. Underwood S, Lyratzopoulos G, Saunders CL. Breast, prostate, colorectal, and lung cancer incidence and risk factors in women who have sex with women and men who have sex with men: a cross-sectional and longitudinal analysis using UK Biobank. Cancers (Basel) 2023;15(7):2031 [FREE Full text] [doi: 10.3390/cancers15072031] [Medline: 37046692]
54. Pitman A, Marston L, Lewis G, Semlyen J, McManus S, King M. The mental health of lesbian, gay, and bisexual adults compared with heterosexual adults: results of two nationally representative English household probability samples. Psychol Med 2021:1-10 [FREE Full text] [doi: 10.1017/S0033291721000052] [Medline: 33592165]
55. Crowe S, Barker E, Roberts M, Lloyd L, de Barros CM, Rebelo-Harris B, et al. Are we asking the right questions? Working with the LGBTQ+ community to prioritise healthcare research themes. Res Involv Engagem 2021;7(1):64 [FREE Full text] [doi: 10.1186/s40900-021-00298-7] [Medline: 34556178]
56. MacCarthy S, Saunders CL, Elliott MN. Sexual minority adults in England have greater odds of chronic mental health problems: variation by sexual orientation, age, ethnicity, and socioeconomic status. LGBT Health 2022;9(1):54-62 [FREE Full text] [doi: 10.1089/lgbt.2021.0011] [Medline: 34882021]

57. Bécares L, Kneale D. Inequalities in mental health, self-rated health, and social support among sexual minority young adults during the COVID-19 pandemic: analyses from the UK millennium cohort study. Soc Psychiatry Psychiatr Epidemiol 2022;57(10):1979-1986 [FREE Full text] [doi: 10.1007/s00127-022-02291-1] [Medline: 35507033]
58. Saunders CL, Flynn S, Massou E, Lyratzopoulos G, Abel G, Burt J. Sociodemographic inequalities in patients' experiences of primary care: an analysis of the general practice patient survey in England between 2011 and 2017. J Health Serv Res Policy 2021;26(3):198-207 [FREE Full text] [doi: 10.1177/1355819620986814] [Medline: 33517786]
59. Saunders CL, Gkousis E. Impact of telephone triage on access to primary care for people living with multiple long-term health conditions: rapid evaluation. Health Soc Care Deliv Res 2022;10(18):1-48 [FREE Full text] [doi: 10.3310/ucce5549] [Medline: 35793423]
60. Booker CL, Meads C. Sexual orientation and the incidence of COVID-19: evidence from understanding society in the UK longitudinal household study. Healthcare (Basel) 2021;9(8):937 [FREE Full text] [doi: 10.3390/healthcare9080937] [Medline: 34442074]
61. Goldacre B, Morley J. Better, broader, safer: using health data for research and analysis. A Review Commissioned by the Secretary of State for Health and Social Care. 2022. URL: https://assets.publishing.service.gov.uk/media/624ea0ade90e072a014d508a/goldacre-review-using-health-data-for-research-and-analysis.pdf [accessed 2024-04-10]
62. Ford T, Mansfield KL, Markham S, McManus S, John A, O'Reilly D, et al. The challenges and opportunities of mental health data sharing in the UK. Lancet Digit Health 2021;3(6):e333-e336 [FREE Full text] [doi: 10.1016/S2589-7500(21)00078-9] [Medline: 34044999]
63. Romanelli RJ, Picken N, Adams A. Counting LGBTQ+ Lives in England and Wales: A Critical Reflection on Emerging Findings from the 2021 Census. Santa Monica, CA: RAND Corporation; 2023.
64. Braybrook D, Bristowe K, Timmins L, Roach A, Day E, Clift P, et al. Communication about sexual orientation and gender between clinicians, LGBT+ people facing serious illness and their significant others: a qualitative interview study of experiences, preferences and recommendations. BMJ Qual Saf 2023;32(2):109-120 [FREE Full text] [doi: 10.1136/bmjqs-2022-014792] [Medline: 36657773]
65. van Daalen KR, Davey F, Norman C, Ford JA. Health equity audits: a systematic review of the effectiveness. BMJ Open 2021;11(11):e053392 [FREE Full text] [doi: 10.1136/bmjopen-2021-053392] [Medline: 34764176]
66. Participate in research. LGBT Foundation. URL: https://lgbt.foundation/research/participate [accessed 2024-04-10]
67. Plsek PE, Greenhalgh T. Complexity science: the challenge of complexity in health care. BMJ 2001;323(7313):625-628 [FREE Full text] [doi: 10.1136/bmj.323.7313.625] [Medline: 11557716]
68. Evidence review: gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria. National Institute for Health and Care Excellence. 2020. URL: https://cass.independent-review.uk/nice-evidence-reviews/ [accessed 2024-04-10]
69. Evidence review: gender-affirming hormones for children and adolescents with gender dysphoria. National Institute for Health and Care Excellence. 2020. URL: https://cass.independent-review.uk/wp-content/uploads/2022/09/20220726_Evidence-review_Gender-affirming-hormones_For-upload_Final.pdf [accessed 2024-04-10]
70. Ethical research: good practice guide to researching LGBT communities and issues. LGBT Foundation. URL: https://dxfy8lrzbpywr.cloudfront.net/Files/1a884870-453a-429d-a213-399a9502472c/Ethics%2520Guide.pdf [accessed 2024-04-10]
71. Cowan K. INVOLVE: A practical guide to being inclusive in public involvement in health research - lessons from the reaching out programme. National Institute for Health Research. 2020. URL: https://www.invo.org.uk/wp-content/uploads/2020/03/INVOLVE_RO_report_FINAL_180220.pdf [accessed 2024-04-10]
72. de Barros CM, Saunders K, Crowe S. LGBTQ+ research and involvement – what have we learnt? National Institute for Health and Care Research: School for Primary Care Research. 2022. URL: https://www.spcr.nihr.ac.uk/news/blog/lgbtq-research-and-involvement-2013-what-have-we-learnt, [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

*Edited by T de Azevedo Cardoso; submitted 03.10.23; peer-reviewed by C Li, E Fedele; comments to author 16.02.24; revised version received 28.02.24; accepted 26.03.24; published 01.05.24.*

*Please cite as:*
*Saunders CL*
*Using Routine Data to Improve Lesbian, Gay, Bisexual, and Transgender Health*
*Interact J Med Res 2024;13:e53311*
*URL: https://www.i-jmr.org/2024/1/e53311*
*doi:10.2196/53311*
*PMID:38691398*

Viewpoint

# Ad Hoc Modifications to a High Dependency Psychiatric Unit for People With Dementia During the COVID-19 Period

Thanita Pilunthanakul[1*], BSc, MA, MD; Giles Ming Yee Tan[2*], MBBS, M Med

[1]MOH Holdings Pte Ltd, Singapore, Singapore

[2]Institute 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.

***(Interact J Med Res 2024;13:e49618)*** doi:10.2196/49618

**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).

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

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

## *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

1. Lim WS, Soh KC, Surej J, Goh KH. Examining the role of a psychiatry department in Singapore amidst COVID-19. Asian J Psychiatr 2020 Dec;54:102296 [FREE Full text] [doi: 10.1016/j.ajp.2020.102296] [Medline: 32682304]

2. Poremski D, Subner SH, Lam GF, Dev R, Mok YM, Chua HC, et al. Effective infection prevention and control strategies in a large, accredited, psychiatric facility in Singapore. Infect Control Hosp Epidemiol 2020 Oct 23;41(10):1238-1240 [FREE Full text] [doi: 10.1017/ice.2020.163] [Medline: 32321623]
3. Wang H, Li T, Barbarino P, Gauthier S, Brodaty H, Molinuevo JL, et al. Dementia care during COVID-19. Lancet 2020 Apr 11;395(10231):1190-1191 [FREE Full text] [doi: 10.1016/S0140-6736(20)30755-8] [Medline: 32240625]
4. D'Cruz M, Banerjee D. Caring for persons living with dementia during the COVID-19 pandemic: advocacy perspectives from India. Front Psychiatry 2020 Oct 27;11:603231 [FREE Full text] [doi: 10.3389/fpsyt.2020.603231] [Medline: 33192744]
5. Cohen G, Russo MJ, Campos JA, Allegri RF. Living with dementia: increased level of caregiver stress in times of COVID-19. Int Psychogeriatr 2020 Nov;32(11):1377-1381 [FREE Full text] [doi: 10.1017/S1041610220001593] [Medline: 32729446]
6. Brown EE, Kumar S, Rajji TK, Pollock BG, Mulsant BH. Anticipating and mitigating the impact of the COVID-19 pandemic on Alzheimer's disease and related dementias. Am J Geriatr Psychiatry 2020 Jul;28(7):712-721 [FREE Full text] [doi: 10.1016/j.jagp.2020.04.010] [Medline: 32331845]
7. Waller S, Masterson A. Designing dementia-friendly hospital environments. Future Hosp J 2015 Feb;2(1):63-68 [FREE Full text] [doi: 10.7861/futurehosp.2-1-63] [Medline: 31098081]
8. Garland K, Beer E, Eppingstall B, O'Connor DW. A comparison of two treatments of agitated behavior in nursing home residents with dementia: simulated family presence and preferred music. Am J Geriatr Psychiatry 2007 Jun;15(6):514-521. [doi: 10.1097/01.JGP.0000249388.37080.b4] [Medline: 17293386]
9. Cohen-Mansfield J, Marx MS, Dakheel-Ali M, Regier NG, Thein K, Freedman L. Can agitated behavior of nursing home residents with dementia be prevented with the use of standardized stimuli? J Am Geriatr Soc 2010 Aug 03;58(8):1459-1464 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.02951.x] [Medline: 20579167]
10. Hicks-Moore SL. Relaxing music at mealtime in nursing homes: effects on agitated patients with dementia. J Gerontol Nurs 2005 Dec;31(12):26-32. [doi: 10.3928/0098-9134-20051201-07] [Medline: 16375095]

## Abbreviations

**HDPCU:** High Dependency Psychiatric Unit
**IMH:** Institute of Mental Health

*Edited by T de Azevedo Cardoso; submitted 03.06.23; peer-reviewed by HY Goh, V Rashedi; comments to author 12.12.23; revised version received 22.01.24; accepted 13.05.24; published 11.06.24.*

*Please cite as:*
*Pilunthanakul T, Tan GMY*
*Ad Hoc Modifications to a High Dependency Psychiatric Unit for People With Dementia During the COVID-19 Period*
*Interact J Med Res 2024;13:e49618*
*URL: https://www.i-jmr.org/2024/1/e49618*
*doi:10.2196/49618*
*PMID:38861715*

Viewpoint

# Resilience Informatics: Role of Informatics in Enabling and Promoting Public Health Resilience to Pandemics, Climate Change, and Other Stressors

M Sriram Iyengar[1], BTECH, MSc, MS, PhD; Maiya G Block Ngaybe[2], MPH; Myla Gonzalez[2]; Mona Arora[2], MSPH, PhD

[1]University of Arizona College of Medicine, Phoenix, AZ, United States

[2]Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

**Corresponding Author:**
M Sriram Iyengar, BTECH, MSc, MS, PhD
University of Arizona College of Medicine
475 North 5th Street
E967C
Phoenix, AZ, 85004
United States
Phone: 1 2817934733
Email: msiyengar@arizona.edu

## Abstract

Climate change, local epidemics, future pandemics, and forced displacements pose significant public health threats worldwide. To cope successfully, people and communities are faced with the challenging task of developing resilience to these stressors. Our viewpoint is that the powerful capabilities of modern informatics technologies including artificial intelligence, biomedical and environmental sensors, augmented or virtual reality, data science, and other digital hardware or software, have great potential to promote, sustain, and support resilience in people and communities. However, there is no "one size fits all" solution for resilience. Solutions must match the specific effects of the stressor, cultural dimensions, social determinants of health, technology infrastructure, and many other factors.

***(Interact J Med Res 2024;13:e54687)*** doi:10.2196/54687

**KEYWORDS**

health informatics; data science; climate change; pandemics; COVID-19; migrations; mobile phone

## Introduction

In recent years, a series of stressful events has caused substantial economic, physical, and medical harm worldwide. As of June 2024, the COVID-19 pandemic that began in early 2020 has been estimated to be responsible for about 1,193,535 COVID-19–related deaths in the United States alone [1]. It was responsible for one of the worst global recessions in recent history in addition to causing great disruption to normal life due to lockdowns, closures of educational institutions, shutting down of tourism and travel, and a host of other adverse impacts [2-6].

In addition to pandemics, climate change has emerged as an economic and public health threat [7-10]. Climate change continues to intensify the emergence of pathogenic diseases and exacerbate pandemics and infectious disease outbreaks such as Ebola, COVID-19, and Mpox [11,12]. Mental health has been identified as a casualty of climate change worldwide [13-16]. Migrant and refugee populations are also subject to increased health risks [17,18].

Adverse events, whether disease outbreaks or climate-driven events, cause significant stress to people, communities, and organizations, threatening human health and animal health while exacerbating systemic inequities. These events highlight the importance of building resilience in communities to prepare for and mitigate the impact of these stressors on people and communities.

Resilience has been defined in several ways (see Multimedia Appendix 1 [19-22]). In this paper, we will use the comprehensive definition from the US Agency for International Development: "the ability of people, households, communities, systems, and countries to mitigate, adapt to, and recover from shocks and stresses in a manner that reduces age and chronic vulnerabilities and facilitates equitable health outcomes" [19]. The discipline of resilience research investigates and develops

evidence-based strategies, technologies, knowledge, and information to build preparedness and inform response and recovery across communities [23]. These stressors can include pandemics, floods, droughts, and displacement of populations due to conflicts, among others. It is noteworthy that these events can be recurrent [24]. While these stressors can affect all aspects of society and the environment, including agriculture, economics, industry, and infrastructure, in this paper, we will focus on resilience relating to public health threats. Thus, all references to resilience in the following should be understood as relating to health.

Informatics and communications technologies are essential infrastructures worldwide. Due to their increasingly powerful capabilities, these technologies have great potential to play important roles in supporting resilience. In this paper we propose Resilience Informatics (RI) as the application of informatics techniques to materially improve and promote the ability of people, communities, and organizations, to effectively cope with natural and man-made stressors [25]. We introduce RI as a people-centric field of study, research, and development.

### Climate and Pandemic Resilience

Climate change threatens human health and well-being in myriad ways: warmer temperatures are expected to increase the risk of vector-borne diseases [26]; elevated temperatures are projected to result in deaths of millions by the year 2100 [27]. Of particular concern among certain island nations and coastal zone communities is the rise in sea levels which could result in the submergence of these locations. This could lead to large-scale displacement of people [28,29]. The impacts of a changing climate are already evident across the United States: in 2012 alone, the economic cost of climate change effects including wildfires, ozone pollution, heat waves, mosquito-borne disease was US $10 billion (in 2018 dollars) [30]. While direct impacts on health include morbidity and mortality associated with extreme weather events such as heatwaves and floods, indirect health impacts to adults and children are caused by changes in the dynamics of vector-borne and water-borne diseases, malnutrition due to decreased food security, and population displacement that may arise through alterations in our environmental and social systems [31-33].

Climate resilience refers to the ability to prepare for, recover from, and adapt to the threats associated with a changing climate, including but not limited to more frequent and severe extreme weather events and prolonged droughts. Climate resilience centers around societies and communities mobilizing resources and partners to anticipate these risks, reducing community vulnerability to those risks through infrastructure and other investments, preparing for, responding to, and recovering from these events.

### Crisis Informatics and RI

Crisis informatics, also known as disaster informatics, is “a multidisciplinary field combining computing and social science knowledge of disasters” [34,35]. Crisis informatics has a major focus on the informatics needs of first responders such as firemen, construction workers, and health care workers including medics, nurses, and physicians. Crisis informatics tools have been proposed to improve the efficiency of crisis response methods such as evacuations, provision of medical supplies, and disaster preparation. Mobile tools [36] have been used during crises to assist affected individuals and these can be viewed as examples of RI tools for, typically, short-term immediate responses to crises. By contrast, RI has its sole focus on improving the ability of people and communities to successfully cope not only with disasters but also to prepare for long-lasting and recurrent threats to health.

## About Resilience Informatics

In the broad sense, RI encompasses (1) data science and artificial intelligence to aid the design, development, and evaluation of resilience strategies; and (2) hardware, software, and systems that translate resilience strategies into customized tools for people and communities. Well-designed RI tools can play an important role in strengthening local capacities in public health. Measurement of health resilience among people and communities is of particular importance. The recent COVID-19 pandemic saw the development and deployment of a host of informatics tools including contact tracing, disease surveillance, and messaging [37-41].

### Classification of Stressors

The nature, design, requirements, and development processes of RI tools could vary across 2 types of stressor events as defined below. We identify drivers of resilience whose impact could be enhanced by RI tools and systems.

#### *Type (1) Acute Events (eg, Floods or Wildfires)*

Here, a major role of RI is to develop tools and systems that disseminate educational information on resources for resilience and recovery and gather data for computing and measuring resilience metrics. If a cellular data infrastructure is available, as in crises that do not affect basic infrastructures, the widespread availability of cell phones and the very high penetration rates of smartphones can provide very useful venues for this purpose [42]. Data gathered from cell phones can provide geo-coded and time-stamped information enabling dynamic identification of “hot spots” where resource allocation can be made to the most affected areas. These data can also be used to develop management dashboards for the benefit of public officials to monitor and inform decision-making. Typically, RI tools to promote resilience in type 1 events need to be developed rapidly, as soon as possible after the event has occurred to provide speedy assistance such as guidelines for recovery to affected persons.

#### *Type (2) Long-Term “Chronic,” Recurrent, and Persistent Stressors*

Long-term and persistent effects of climate change, including hydrometeorological events such as droughts and warming temperatures, fall into this category. Another example is the transition of pandemics from an acute phase in which new infections occur at high rates, into a long-term phase in which new infections occur at greatly reduced rates while a percentage of persons who were previously infected have persistent consequences. The COVID-19 pandemic is an example [43]. Chronic stressors can cause acute stressors, for example,

excessive rainfall caused by changes in climate patterns. The major need for resilience to chronic stressors is to develop a culture of resilience. Here, data science and machine learning can identify data patterns that contribute to the long-term stressor as well as generate insights about efficacious responses and develop tools to reduce inequitable responses. Informatics tools developed for acute stressor response could be transitioned into systems to enhance resilience for chronic stressors. As noted previously, even in low- and middle-income countries (LMICs), cell phones have achieved very high penetration, providing a cost-effective and highly accessible means for providing targeted information and gathering data. As an example, in the last 10 years, mobile health tools in LMICs [44-47] have been intensively researched and could become common practice [36,46-50]. Notably, in LMICs the WhatsApp system is often used for videoconferencing between clinicians and remote patients [51].

## Behavior Change in Resilience and the Role of Persuasive Technology in RI

Responding to stressors can require people and communities to change their behavior. A vivid illustration occurred during the COVID-19 pandemic. People worldwide needed to change their normal behavior both on individual and social levels. Frequent handwashing and the use of sanitizing wipes were recommended. Mask-wearing was mandatory when doing commonplace activities such as shopping and traveling on public transport.

Persuasive Technology (PT) [52] is concerned with the design of noncoercive tools and technologies to change human behavior. Following the original work of Fogg [52,53], Oinas-Kukkonen and Harjumaa [54] developed the Persuasive Systems Design (PSD) framework to guide the development of PT tools. Apart from Fogg 7 primary task support postulates, the PSD framework includes 3 major components in the design of PT tools: dialog support, system credibility support, and social support. A major focus in recent years has been the application of PT to support healthful behavior change [55-57]. Since, as noted above, behavior change can be an important comment on developing resilience, PT and PSD could help guide the design of effective RI tools and systems.

## Drivers of Resilience and Informatics

Drivers of resilience include social contexts, community factors, economics, institutions, and infrastructure. An important area of future research is to identify the role of RI and the designs of RI systems to enhance the effectiveness of these drivers of resilience.

### *Social Contexts*

Societal structures tend to be either individualistic or communal in nature and shape RI systems developed for the given level of structure (town, county, or state) [58-61]. In the example provided above, social media is highlighted as a communication mechanism. The use of social media as an adaptive management tool has become increasingly prevalent which can be seen through the move of the news to social media sites [62] and the use of technology for this purpose has been seen through the development and use of messaging apps for flash flood warnings, silver alerts, and amber alerts [63,64]. These systems have proved effective in communities, and so the adaptive management approach and systems, which can easily change to include more information and updates about the status of the crisis, emergency, or disaster, can contribute substantially to developing a substantial resilience procedure.

### *Community Factors*

Community factors that drive resilience include social groups, religious charitable organizations, and community food banks [61]. During the first part of the COVID-19 pandemic in 2020, communities in New York worked to form social groups through social media (Facebook or WhatsApp chats), which worked to establish a mutual aid system to provide resources such as groceries, masks, etc, as well as services such as child care, pet care, and running errands, for at-risk individuals in the community [65]. These mutual aid groups drove resilience within their communities through a demonstrated dedication to the well-being of the population.

### *Institutions and Infrastructure Factors*

The measures that an institution may take in approaching a stressor, and the adaptability of those measures, are determining factors in the effectiveness of resilience-driven responses to crises. During the COVID-19 pandemic, governments had to quickly develop responses to the pandemic that would prevent the spread of the disease and reduce its prevalence [66]. Hong Kong, Singapore, and Japan maintained contact throughout the pandemic to implement systems restricting the travel of their citizens, which would in turn contain the spread of COVID-19. Additionally, each of these governments developed systems of communication between the health care providers and the government to ensure the practiced pandemic response was maximally effective at any given time [67]. These responses to the pandemic, as aforementioned, were developed at the time of the pandemic. Institutions establishing RI systems must be able to develop adaptive systems that can rapidly respond to changes in the conditions of the pandemic allowing for an effective resiliency response to be deployed.

## Development of RI Systems

Systems development can be greatly aided by following a conceptual framework. Apart from systematic development and evaluation, the benefits of doing so can include flexibility, extensibility (the ability to add features easily), scalability (the ability of the system to be used by greater numbers of people without redevelopment), and others [68]. We propose the following 6-component framework (Textbox 1) [25] as a guide for efficient and effective development of RI systems.

**Textbox 1.** The 6-component development framework for Resilience Informatics (RI) tools and systems.

**Component 1: Team**

- The team should be multidisciplinary and multisectoral, including experts in the target population, the environment, the technologies, and the specific resilience context.

**Component 2: Requirements**

- Requirements may include system features that maximize system effectiveness, for example, multilanguage capability, consistent appearance, and functionality across operating systems (eg, Android and iOS); screen sizes; judicious use of multimedia.

**Component 3: Information**

- The system must provide the correct and most pertinent information at the right spatial and temporal scales.

**Component 4: Design considerations**

- Design of an RI system, including the user interface, must consider aspects including but not limited to the target population's characteristics, social determinants of health, and cultural factors.

**Component 5: Implementation**

- Efficient and rapid implementation strategies need to be developed. The project must have a process for responding to the changing needs in the targeted population.

**Component 6: Evaluation**

- The system should integrate an evaluation strategy to continuously assess the impact of the project on achieving its objectives.

### *Component 1: Team*

It is important to recruit a multidisciplinary, multi-sectoral team that represents all aspects or factors that contribute to resilience.

### *Component 2: Requirements*

Evolving project scope and changes in the operating environment as the effects of a stressor unfold suggest that RI systems requirements need to be flexible to accommodate these changes. An adaptive management approach including the Agile development model [69] would be most appropriate. Ideally, the development methodology can enable rapid changes to the system even doing implementation and during subsequent operations to accommodate changing requirements. Requirements gathering may take a lot of time but will likely have to be accelerated in the case of acute events to develop the system and make it available to the target users as soon as possible.

### *Component 3: Information*

RI interventions are based on data that help to provide correct and relevant information. This information must be tailored to the target users and match the cultural, linguistic, literacy, educational, and economic status of the target population.

### *Component 4: Design Considerations*

The design of an RI intervention must be tailored to the characteristics of the target population. These include the intended level of resilience, that is, the household, the community, and the health care system. The system must be designed to match the technological capabilities available in the target context. For example, the availability of clean and reliable electricity can be a constraint in some locations. In addition, the intervention must match the education, culture, language, and other characteristics of target users. To aid engagement and sustained benefits, frameworks for technology adoption and principles of PT and PSD could be useful [70]. Integration of theory in the design of the RI project is another important consideration that would occur at this stage to support its effectiveness and impact.

### *Component 5: Implementation*

The RI project or system must have an efficient and rapid implementation using an adaptive management approach and considering currently available resources (eg, availability of data and communications infrastructures). The process should provide maximum benefits while incorporating a feedback loop to respond to changes in the operating environment that may necessitate changing requirements.

### *Component 6: Evaluation*

Finally, it is important to integrate a logic model or similar implementation-guiding tool into the design of the RI system or project to support continuous evaluation efforts that can inform needed changes and modifications.

## *RI Case Study: AZCOVIDTXT*

As an illustration of the RI development framework, we present the following case study of a technology solution implemented to support the resilience of the people of the state of Arizona in the United States to the COVID-19 pandemic [71].

### Context

Beginning in March 2020, the COVID-19 pandemic caused immense disruptions to normal life. These disruptions led to an atmosphere of fear in which uncertainty, rumors, and misinformation caused widespread public confusion preventing people from responding effectively.

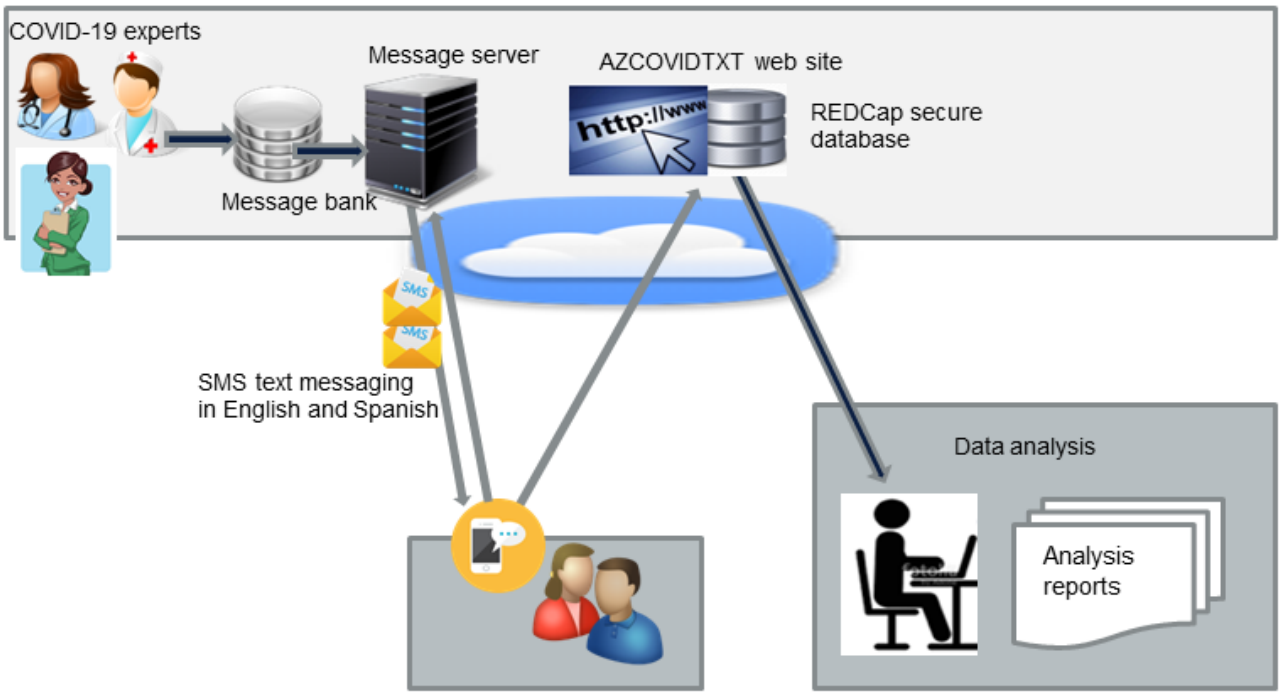
## Purpose

To help alleviate this situation and enhance resilience among Arizonans, faculty including some of the authors of this paper, at the University of Arizona identified the need for an informatics strategy to provide much-needed authoritative and timely COVID-19 information to the public, and to gather information on how the pandemic was impacting communities' health, well-being, and them financially.

## Informatics Solution

We chose text messaging, also known as SMS, as the messaging modality because it is low-footprint, ubiquitous, and eliminates the time and expense of multi-language app development for iOS and Android. SMS text messaging is an established multi–operating system technology, shown to be effective in health care [72,73], usage of text messaging is universal, and no specific training is needed for individuals to read and respond to text messages. SMS text messaging is inexpensive, and unlimited text messaging is often included in many cell phone plans. Enrollment occurred by text messaging a toll-free number, and via the AZCOVIDTXT website during which enrollees indicate their preference for receiving messages or surveys in English or Spanish. Enrollment was limited to those providing Arizona zip codes. For data collection we developed a simple REDCap (Research Electronic Data Capture; Vanderbilt University)-based [74] mobile survey tool (Figure 1).

**Figure 1.** AZCOVIDTXT system. CDC: Centers for Disease Control and Prevention; REDCap: Research Electronic Data Capture.




## System Design Considerations

Requirements identified included supporting individuals in both English and Spanish, ease of use and installation, and the capability to gather data on COVID-19–related challenges. A mobile health approach was identified as the optimal strategy since even among disadvantaged groups, cell phone access or ownership is at 100% and smartphone access is at least 85% [75]. The resulting system, called AZCOVIDTXT, was developed and deployed in 4 weeks beginning in April 2020 and operated until March 2022.

Messages and other content were developed and curated by specialists in public health, health behavior change, and health communications, assisted by graduate students and staff at the University of Arizona. The message content was derived from authoritative sources such as the Centers for Disease Control, and new messages were developed weekly to reflect the latest information, including countering misinformation, on the availability of vaccines, outbreaks of SARS-CoV-2 variants, and other evolving events.

By March 2022, a total of 3746 households from 225 Arizona zip codes were enrolled in AZCOVIDTXT, and more than 522,000 text messages providing COVID-19–related information had been sent. Except for a few outages, surveys were sent about every 10 days and messages 3 times a week. Curated content consisting of nearly 200 news updates, SMS text messages to enrollees, and social media posts (@AZCOVIDTXT on Instagram [Instagram from Meta], Facebook [Meta], and Twitter [X Corp]) had been developed.

AZCOVIDTXT applied the 6-components principles for RI (Textbox 2) and successfully developed and deployed the system.

Due to the waning of the COVID-19 pandemic in early 2023, this system was discontinued.

**Textbox 2.** Application of 6-component Resilience Informatics development framework in AZCOVIDTXT.

**Component 1: Team**

- A multidisciplinary team of experts in informatics, mobile health, infectious disease epidemiology, health promotions, health behavior change, programmers, and students was assembled rapidly.

**Component 2: Requirements**

- Simplicity and ease of use were primary requirements. Messaging should be in English and Spanish and received on cell phones due to their very widespread use and minimal cognitive demand on a user. These requirements precluded the use of apps because apps take time to develop, need maintenance, and can require the user to download them on their phones. SMS text messaging in English and Spanish was selected since it is available on all types of phones (not just smartphones); the user interface is understood universally.

**Component 3: Information**

- Messages and content were sourced from authoritative sources (Centers for Disease Control and Prevention, World Health Organization) and curated by an infectious disease epidemiologist and health promotions expert assisted by University of Arizona graduate students and staff [71]. Weekly messages reflected the latest information and conditions (ie, vaccine availability or SARS-CoV-2 variant outbreaks).

**Component 4: Design**

- The team applied the requirements of simplicity and ease of use to design the AZCOVIDTXT website.

**Component 5: Implementation**

- The system was developed and deployed in 4 weeks using public-domain software and systems such as REDCap (Research Electronic Data Capture). This included a web site (now discontinued), and toll-free numbers for enrollment.

**Component 6: Evaluation**

- The performance of the system was evaluated by the number of families enrolled and the reach of the system.

# Discussion, Issues, and Limitations

## Overview

The public health system comprised of a web of federal, state, and local agencies, hospitals, nonprofit agencies, and businesses is at the forefront of the health response to climate change. Building climate resilience for the public health system requires understanding the complexity of the climate, health, and human systems that have unique innate behaviors and structures that contribute to risks and vulnerabilities and need to be considered when informing any community resilience-building strategy. The development of resilience could benefit from a systems-wide informatics approach that engages partners and stakeholders in the processes involved in recognizing threats, determining capacity, informing solutions, recovering from a crisis, and adapting the process to enhance our capacity to deal with the next crisis. RI systems could also play an important supporting role in the US Agency for International Development's Program Cycle [76] for enhancing resilience.

## Issues and Limitations

It is important to note that RI tools and systems are subject to operating limitations caused by the stressor event. For example, issues such as inaccessibility to sufficient cellular network connection, low health literacy, and fragility of cellular networks may interfere with the effectiveness of an RI project. Older adult populations, who often tend to be largely represented in rural areas, may not be able to access new technology easily due to the rapid development of new technologies. Additionally, cellular networks may be overused in the case of an emergency which may lead to issues with communication via technology. Initiatives, such as FirstNet, are being put into place to try to address this issue, especially for first responders and others who need to react quickly and coordinate with their teams during an emergency, but the public at large may still be negatively impacted [77]. Additionally, it is important to recognize the limitations of informatics tools being able to support and enhance resiliency. Resilience is a very large issue that requires a multilevel and multifaceted response, and informatics is only one part of the larger response that would need to occur for a community to be able to cope effectively in the face of a disaster.

## Equity and Inclusion

The current issues and limitations of RI data communication lie within already existing socioeconomic discrimination and access difficulty of the methods of communication. Yang et al [78] discuss how underserved communities, identified as minorities, older adults, and the poor, were not considered in deciding systems for spreading disaster informatics data. By effect, these groups experience sizable differences in the disaster relief provided to them. As RI systems continue to become more heavily technology-reliant, these gaps in treatment could be further emphasized. Virapongse et al [79] detail how these gaps are becoming ever more important to address, as communities become more reliant on the data provided to them. They argue that working directly with stakeholders in underserved communities would allow for their issues to be directly addressed and prevent further inequitable treatment of these groups.

Tailored tools, interventions, and programs are a critical aspect of informing resilience and RI allows for this strategic, locally relevant decision-making. No model for RI will be effective for

every community, as the conventions and resources of each area differ greatly. Therefore, RI system designs must allow decision makers to adapt the model to fit the needs of their area. For information distribution, this would mean that decision makers would adapt their RI systems to account for the technology available to the given community and the preferences of each group of people in how they desire to receive information.

## Conclusions

RI has great potential to enable and support strategies, resources, and technologies that enhance the ability of individuals, populations, and the environment to respond successfully to natural and man-made stressors. The conditions under which RI systems function impose design and other constraints that make RI distinct from other informatics environments. There is a great need for research to establish the basic principles of RI leading to the efficient development of RI systems to support resilience in the United States and globally. The 6-component framework presented here could be a useful guide to the efficient development and effective deployment of informatics tools to promote resilience in public health.

## Acknowledgments

This research was supported in part by a grant from the Arizona Technology, Research and Innovation Fund. The authors would like to gratefully thank Professors Sharon Collinge, Zack Guido, Kacey Ernst, Nirav Merchant, and Michael Fallon, all of the University of Arizona, for support, encouragement and many helpful discussions.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of resilience.
[DOCX File , 14 KB - ijmr_v13i1e54687_app1.docx ]

## References

1. Trends in United States COVID-19 deaths, emergency department (ED) visits, and test positivity by geographic area. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker/#trends_totaldeaths_select_00 [accessed 2024-07-11]
2. Grover S, Sahoo S, Mehra A, Avasthi A, Tripathi A, Subramanyan A, et al. Psychological impact of COVID-19 lockdown: an online survey from India. Indian J Psychiatry 2020;62(4):354-362 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_427_20] [Medline: 33165368]
3. Bennett G, Young E, Butler I, Coe S. The impact of lockdown during the COVID-19 outbreak on dietary habits in various population groups: a scoping review. Front Nutr 2021;8:626432 [FREE Full text] [doi: 10.3389/fnut.2021.626432] [Medline: 33748175]
4. Pokhrel S, Chhetri R. A literature review on impact of COVID-19 pandemic on teaching and learning. Higher Educ Future 2021;8(1):133-141. [doi: 10.1177/2347631120983481]
5. Mofijur M, Fattah IMR, Alam MA, Islam ABMS, Ong HC, Rahman SMA, et al. Impact of COVID-19 on the social, economic, environmental and energy domains: lessons learnt from a global pandemic. Sustain Prod Consum 2021;26:343-359. [doi: 10.1016/j.spc.2020.10.016] [Medline: 33072833]
6. Koh HK, Geller AC, VanderWeele TJ. Deaths from COVID-19. JAMA 2021;325(2):133-134. [doi: 10.1001/jama.2020.25381] [Medline: 33331884]
7. Tol RSJ. The economic impacts of climate change. Rev Environ Econ Policy 2018;12(1):4-25. [doi: 10.1093/reep/rex027]
8. Romanello M, Di Napoli C, Drummond P, Green C, Kennard H, Lampard P, et al. The 2022 report of the Lancet Countdown on health and climate change: health at the mercy of fossil fuels. Lancet 2022;400(10363):1619-1654. [doi: 10.1016/S0140-6736(22)01540-9] [Medline: 36306815]
9. Haines A, Kovats RS, Campbell-Lendrum D, Corvalan C. Climate change and human health: impacts, vulnerability and public health. Public Health 2006;120(7):585-596. [doi: 10.1016/j.puhe.2006.01.002] [Medline: 16542689]
10. Fourth national climate assessment. USGCRP. Washington, DC: U.S. Global Change Research Program; 2018. URL: https://nca2018.globalchange.gov [accessed 2023-11-16]
11. Mora C, McKenzie T, Gaw IM, Dean JM, von Hammerstein H, Knudson TA, et al. Over half of known human pathogenic diseases can be aggravated by climate change. Nat Clim Chang 2022;12(9):869-875 [FREE Full text] [doi: 10.1038/s41558-022-01426-1] [Medline: 35968032]
12. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. Nat Rev Microbiol 2022;20(4):193-205 [FREE Full text] [doi: 10.1038/s41579-021-00639-z] [Medline: 34646006]
13. Clayton S. Climate change and mental health. Curr Environ Health Rep 2021;8(1):1-6. [doi: 10.1007/s40572-020-00303-3] [Medline: 33389625]

14. Clayton S. Climate anxiety: psychological responses to climate change. J Anxiety Disord 2020;74:102263. [doi: 10.1016/j.janxdis.2020.102263] [Medline: 32623280]
15. Corvalan C, Gray B, Prats EV, Sena A, Hanna F, Campbell-Lendrum D. Mental health and the global climate crisis. Epidemiol Psychiatr Sci 2022;31:e86 [FREE Full text] [doi: 10.1017/S2045796022000361] [Medline: 36459133]
16. Walinski A, Sander J, Gerlinger G, Clemens V, Meyer-Lindenberg A, Heinz A. The effects of climate change on mental health. Dtsch Arztebl Int 2023;120(8):117-124 [FREE Full text] [doi: 10.3238/arztebl.m2022.0403] [Medline: 36647584]
17. World Health Organization. Common health needs of refugees and migrants: literature review. Geneva: World Health Organization; 2021. URL: https://iris.who.int/bitstream/handle/10665/346743/9789240033108-eng.pdf?sequence=1 [accessed 2024-07-18]
18. Matlin SA, Depoux A, Schutte S, Flahault A, Saso L. Migrants' and refugees' health: towards an agenda of solutions. Public Health Rev 2018 Dec;39(27):1-55. [doi: 10.1186/s40985-018-0104-9]
19. Resilience. URL: https://www.resiliencelinks.org/about/what-is-resilience [accessed 2024-06-23]
20. Disaster Resilience: A National Imperative. Washington, DC: The National Academies Press; 2012. URL: https://nap.nationalacademies.org/read/13457/chapter/1 [accessed 2024-07-03]
21. Building private-public partnerships. FEMA. Missouri: US Chamber of Commerce Foundation; 2021. URL: https://www.fema.gov/sites/default/files/documents/fema_building-private-public-partnerships.pdf [accessed 2024-07-03]
22. Dokken D. Summary for policymakers. URL: https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf [accessed 2024-06-27]
23. Werner EE. Resilience research. In: Peters RD, Leadbeater B, McMahon RJ, editors. Resilience in Children, Families, and Communities. California: Springer US; 2005:3-11.
24. Branham L. Overcoming recurring crisis through resilience: an analysis of USAID's definition of resilience. Int Dev Community Environ. 2019. URL: https://commons.clarku.edu/idce_masters_papers/236 [accessed 2024-06-27]
25. Iyengar MS, Merchant N, Ernst K, Rains S, Arora M, Ngaybe MGB, et al. Resilience informatics for public health. Stud Health Technol Inform 2024;310:1276-1280. [doi: 10.3233/SHTI231170] [Medline: 38270020]
26. Caminade C, McIntyre KM, Jones AE. Impact of recent and future climate change on vector‐borne diseases. Ann N Y Acad Sci 2019 Jan;1436(1):157-173. [doi: 10.1111/nyas.13950]
27. Rom WN. Estimated global mortality from present to 2100 from climate change. Int J Humanit Soc Sci Rev 2018 Mar;4(2):1-8.
28. Nicholls RJ, Cazenave A. Sea-level rise and its impact on coastal zones. Science 2010;328(5985):1517-1520. [doi: 10.1126/science.1185782] [Medline: 20558707]
29. Dasgupta S, Laplante B, Meisner C, Wheeler D, Yan J. The impact of sea level rise on developing countries: a comparative analysis. Clim Change 2009;93(3-4):379-388. [doi: 10.1007/s10584-008-9499-5]
30. Limaye VS, Max W, Constible J, Knowlton K. Estimating the health-related costs of 10 climate-sensitive U.S. events during 2012. Geohealth 2019;3(9):245-265 [FREE Full text] [doi: 10.1029/2019GH000202] [Medline: 32159045]
31. Myers SS, Bernstein A. The coming health crisis: indirect health effects of global climate change. F1000 Biol Rep 2011;3:3 [FREE Full text] [doi: 10.3410/B3-3] [Medline: 21399764]
32. Bernstein AS, Myers SS. Climate change and children's health. Curr Opin Pediatr 2011;23(2):221-226. [doi: 10.1097/MOP.0b013e3283444c89] [Medline: 21412082]
33. Watts G, Battarbee RW, Bloomfield JP, Crossman J, Daccache A, Durance I, et al. Climate change and water in the UK – past changes and future prospects. Prog Phys Geogr: Earth Environ 2015;39(1):6-28 [FREE Full text] [doi: 10.1177/0309133314542957]
34. Hagar C. Crisis informatics. In: Encyclopedia of Information Science and Technology. USA: IGI Global; 2015:1350-1358.
35. Palen L, Anderson KM. Crisis informatics—new data for extraordinary times. Science 2016;353(6296):224-225. [doi: 10.1126/science.aag2579] [Medline: 27418492]
36. Tan ML, Prasanna R, Stock K, Hudson-Doyle E, Leonard G, Johnston D. Mobile applications in crisis informatics literature: a systematic review. Int J Disaster Risk Reduct 2017;24:297-311. [doi: 10.1016/j.ijdrr.2017.06.009]
37. Bakken S. Informatics is a critical strategy in combating the COVID-19 pandemic. J Am Med Inform Assoc 2020;27(6):843-844 [FREE Full text] [doi: 10.1093/jamia/ocaa101] [Medline: 32501484]
38. Dagliati A, Malovini A, Tibollo V, Bellazzi R. Health informatics and EHR to support clinical research in the COVID-19 pandemic: an overview. Brief Bioinform 2021;22(2):812-822 [FREE Full text] [doi: 10.1093/bib/bbaa418] [Medline: 33454728]
39. Ye J. The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. JMIR Med Inform 2020;8(7):e19866 [FREE Full text] [doi: 10.2196/19866] [Medline: 32568725]
40. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020;27(6):853-859. [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
41. Massoudi BL, Sobolevskaia D. Keep moving forward: health informatics and information management beyond the COVID-19 pandemic. Yearb Med Inform 2021;30(1):75-83 [FREE Full text] [doi: 10.1055/s-0041-1726499] [Medline: 34479380]
42. Smartphone penetration rates 2022. Statista. URL: https://tinyurl.com/27e8hjx6 [accessed 2023-11-16]

43. Levine RL. Addressing the long-term effects of COVID-19. JAMA 2022;328(9):823-824. [doi: 10.1001/jama.2022.14089] [Medline: 35921084]
44. Hamadeh N, Von Rompaey C, Metreau E, Grace Eapen S. New World Bank country classifications by income level: 2022-2023. World Bank Blogs. URL: https://tinyurl.com/3k8ux9pz [accessed 2022-07-01]
45. McCool J, Dobson R, Whittaker R, Paton C. Mobile health (mHealth) in low- and middle-income countries. Annu Rev Public Health 2022;43:525-539 [FREE Full text] [doi: 10.1146/annurev-publhealth-052620-093850] [Medline: 34648368]
46. Iyengar MS, Chang O, Florez-Arango JF, Taria M, Patel VL. Development and usability of a mobile tool for identification of depression and suicide risk in Fiji. Technol Health Care 2021;29(1):143-153. [doi: 10.3233/THC-202132] [Medline: 32538888]
47. Gautham M, Iyengar MS, Johnson CW. Mobile phone-based clinical guidance for rural health providers in India. Health Inf J 2015;21(4):253-266 [FREE Full text] [doi: 10.1177/1460458214523153] [Medline: 24621929]
48. Iyengar MS, Florez-Arango JF. Decreasing workload among community health workers using interactive, structured, rich-media guidelines on smartphones. Technol Health Care 2013;21(2):113-123. [doi: 10.3233/THC-130713] [Medline: 23510972]
49. Chang O, Patel VL, Iyengar S, May W. Impact of a mobile-based (mHealth) tool to support community health nurses in early identification of depression and suicide risk in Pacific Island Countries. Australas Psychiatry 2021;29(2):200-203. [doi: 10.1177/1039856220956458] [Medline: 32961100]
50. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
51. John B, McCreary C, Roberts A. Smartphone technology for clinical communication in the COVID-19 era: a commentary on the concerning trends in data compliance. Front Digit Health 2022;4:816604 [FREE Full text] [doi: 10.3389/fdgth.2022.816604] [Medline: 35392108]
52. Fogg BJ. Overview of captology. In: Fogg BJ, editor. Persuasive Technology. San Francisco: Morgan Kaufmann; 2003:15-22.
53. Fogg behavior model. Behavior Design Lab. URL: https://behaviordesign.stanford.edu/resources/fogg-behavior-model [accessed 2023-11-16]
54. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24. [doi: 10.17705/1CAIS.02428]
55. Iyengar MS, Oinas-Kukkonen H, Win KT. Persuasive technology in biomedical informatics. J Biomed Inform 2018;85:136-137 [FREE Full text] [doi: 10.1016/j.jbi.2018.07.020] [Medline: 30071315]
56. McGowan A, Sittig S, Bourrie D, Benton R, Iyengar S. The intersection of persuasive system design and personalization in mobile health: statistical evaluation. JMIR mHealth uHealth 2022;10(9):e40576 [FREE Full text] [doi: 10.2196/40576] [Medline: 36103226]
57. Almutairi N, Vlahu-Gjorgievska E, Win KT. Persuasive features for patient engagement through mHealth applications in managing chronic conditions: a systematic literature review and meta-analysis. Inform Health Soc Care 2023;48(3):267-291. [doi: 10.1080/17538157.2023.2165083] [Medline: 36650714]
58. Nagi Y, Sender H, Orcutt M, Fouad F, Burgess RA, Devakumar D. Resilience as a communal concept: understanding adolescent resilience in the context of the Syrian refugee crisis in Bar Elias, Lebanon. J Migr Health 2021;3:100046 [FREE Full text] [doi: 10.1016/j.jmh.2021.100046] [Medline: 34405191]
59. Sousa CA, Haj-Yahia MM, Feldman G, Lee J. Individual and collective dimensions of resilience within political violence. Trauma Violence Abuse 2013;14(3):235-254. [doi: 10.1177/1524838013493520] [Medline: 23793902]
60. Revilla JC, Martín P, de Castro C. The reconstruction of resilience as a social and collective phenomenon: poverty and coping capacity during the economic crisis. Eur Soc 2017;20(1):89-110. [doi: 10.1080/14616696.2017.1346195]
61. Pfefferbaum B, Van Horn RL, Pfefferbaum RL. A conceptual framework to enhance community resilience using social capital. Clin Soc Work J 2017;45(2):102-110. [doi: 10.1007/s10615-015-0556-z]
62. Kümpel AS, Karnowski V, Keyling T. News sharing in social media: a review of current research on news sharing users, content, and networks. Soc Media + Soc 2015;1(2):2056305115610141. [doi: 10.1177/2056305115610141]
63. Wister MA, Hernández-Nolasco JA, Pancardo P, Acosta FD, Jara A. Emergency population warning about floods by social media. 2016 Presented at: 2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS); July 06, 2016; Fukuoka, Japan p. 322-327. [doi: 10.1109/imis.2016.146]
64. Gergerich E, Davis L. Silver alerts: a notification system for communities with missing adults. J Gerontol Soc Work 2017;60(3):232-244. [doi: 10.1080/01634372.2017.1293757] [Medline: 28409711]
65. Soden R, Owen E. Dilemmas in mutual aid: lessons for crisis informatics from an emergent community response to the pandemic. Proc ACM Hum-Comput Interact 2021;5(CSCW2):1-19. [doi: 10.1145/3479862]
66. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (oxford COVID-19 government response tracker). Nat Hum Behav 2021;5(4):529-538. [doi: 10.1038/s41562-021-01079-8] [Medline: 33686204]

67. Legido-Quigley H, Asgari N, Teo YY, Leung GM, Oshitani H, Fukuda K, et al. Are high-performing health systems resilient against the COVID-19 epidemic? Lancet 2020;395(10227):848-850 [FREE Full text] [doi: 10.1016/S0140-6736(20)30551-1] [Medline: 32151326]
68. Schmidt DC, Gokhale A, Natarajan B. Frameworks: why they are important and how to apply them effectively. Queue 2004;2(5):66-75. [doi: 10.1145/1016998.1017005]
69. Isa WARWM, Lokman AM, Aris SRS. Engineering rural informatics using agile user centered design. 2014 Presented at: 2014 2nd International Conference on Information and Communication Technology (ICoICT); 2014 May 28; Bandung, Indonesia p. 367-372. [doi: 10.1109/icoict.2014.6914093]
70. Koul S, Eydgahi A. A systematic review of technology adoption frameworks and their applications. J Technol Manage Innovation 2017;12(4):106-113. [doi: 10.4067/s0718-27242017000400011]
71. Rains SA, Crane TE, Iyengar S, Merchant N, Oxnam M, Sprinkle MM, et al. Community-level health promotion during a pandemic: key considerations for health communication. Health Commun 2020;35(14):1747-1749. [doi: 10.1080/10410236.2020.1837443] [Medline: 33084404]
72. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
73. Willcox JC, Dobson R, Whittaker R. Old-fashioned technology in the era of "Bling": is there a future for text messaging in health care? J Med Internet Res 2019;21(12):e16630 [FREE Full text] [doi: 10.2196/16630] [Medline: 31859678]
74. Wright A. REDCap: a tool for the electronic capture of research data. J Electron Resour Med Libr 2016;13(4):197-201. [doi: 10.1080/15424065.2016.1259026]
75. Mobile Fact Sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-10-30]
76. Program cycle. USAID.gov. URL: https://www.usaid.gov/project-starter/program-cycle [accessed 2023-11-10]
77. Nationwide Broadband For First Responders & Public Safety at FirstNet. URL: https://www.firstnet.com/content/firstnet/en.html [accessed 2023-11-16]
78. Yang Y, Zhang C, Fan C, Mostafavi A, Hu X. Towards fairness-aware disaster informatics: an interdisciplinary perspective. IEEE Access 2020;8:201040-201054. [doi: 10.1109/access.2020.3035714]
79. Virapongse A, Duerr R, Metcalf EC. Knowledge mobilization for community resilience: perspectives from data, informatics, and information science. Sustain Sci 2018;14(4):1161-1171. [doi: 10.1007/s11625-018-0612-z]


## Abbreviations

**LMIC:** low- and middle-income country
**PSD:** Persuasive Systems Design
**PT:** Persuasive Technology
**REDCap:** Research Electronic Data Capture
**RI:** Resilience Informatics


*Edited by T de Azevedo Cardoso; submitted 18.11.23; peer-reviewed by M Raimi, DI Chana, H Shih; comments to author 22.01.24; revised version received 19.02.24; accepted 20.06.24; published 12.08.24.*

*Please cite as:*
*Iyengar MS, Block Ngaybe MG, Gonzalez M, Arora M*
*Resilience Informatics: Role of Informatics in Enabling and Promoting Public Health Resilience to Pandemics, Climate Change, and Other Stressors*
*Interact J Med Res 2024;13:e54687*
*URL: https://www.i-jmr.org/2024/1/e54687*
*doi:10.2196/54687*
*PMID:39133540*

Viewpoint

# Narrowing the Digital Divide: Framework for Creating Telehealth Equity Dashboards

Michael J Luke[1,2,3,4], MD; Sansanee Craig[1,5,6], MD; Suzinne Pak-Gorstein[7], MD, MPH, PhD; Marlíse Arellano[8], MBA; Jessica Zhang[8], MPH; S Margaret Wright[9], MD, MPH; John Chuo[1,6], MD, MS; Philip V Scribano[1,6], DO, MSCE

[1]Children's Hospital of Philadelphia, Philadelphia, PA, United States

[2]Clinical Futures, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[3]Leonard Davis Institute of Health Economics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]PolicyLab, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[5]Department of Biomedical Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, United States

[6]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[7]Department of Pediatrics, School of Medicine, University of Washington, Seattle, WA, United States

[8]Boston Children's Hospital, Boston, MA, United States

[9]Children's Mercy Kansas City, University of Missouri-Kansas City School of Medicine, University of Kansas School of Medicine, Kansas City, MO, United States

**Corresponding Author:**
Michael J Luke, MD
Children's Hospital of Philadelphia
3500 Civic Center Blvd
Philadelphia, PA, 19104
United States
Phone: 1 267 426 6634
Email: lukem1@chop.edu

## Abstract

Telehealth presents both the potential to improve access to care and to widen the digital divide contributing to health care disparities and obliging health care systems to standardize approaches to measure and display telehealth disparities. Based on a literature review and the operational experience of clinicians, informaticists, and researchers in the Supporting Pediatric Research on Outcomes and Utilization of Telehealth (SPROUT)–Clinical and Translational Science Awards (CTSA) Network, we outline a strategic framework for health systems to develop and optimally use a telehealth equity dashboard through a 3-phased approach of (1) defining data sources and key equity-related metrics of interest; (2) designing a dynamic and user-friendly dashboard; and (3) deploying the dashboard to maximize engagement among clinical staff, investigators, and administrators.

***(Interact J Med Res 2024;13:e57435)*** doi:10.2196/57435

**KEYWORDS**

telehealth; equity; dashboard; data; framework; televisit; healthcare; disparity; disparities; clinician; clinicians; informaticist; informaticists; researcher; researchers; pediatric; pediatrics; health system; health systems; dashboards; access to care; data source mapping

## Telehealth Equity

The COVID-19 pandemic catalyzed a surge in telehealth adoption [1,2]. However, disparities in access to and adoption of digital health care persist among Black, Hispanic, public-insured, low-income, and rural populations [3,4]. This “digital divide” risks worsening health disparities in these populations [5]. As such, Crawford and Serhal [6] created the Digital Health Equity Framework (DHEF) to guide the equitable design and implementation of future digital health interventions. The DHEF takes into consideration, how individuals’ sociocultural and economic contexts influence intermediate factors, such as environmental stressors and health behaviors, which then drive the digital determinants of health (eg, acceptability of or access to digital health and digital health literacy) at the root of these disparities.

While health systems can use the DHEF to implement equity-minded telehealth strategies, understanding and bolstering the quality of the digital infrastructure within the communities

they care for are critical steps to ensuring equitable access to telehealth [7]. Unfortunately, digital analytics are still lacking in understanding patterns of use for those underserved by technology infrastructure. Dashboards that showcase key performance indicators in real-time have become valuable tools to track health care access, understand disparities, and apply interventions. Yet, there are no consensus guidelines for the creation of telehealth-specific equity dashboards, which can apply the nuanced considerations for telehealth equity outlined through the DHEF to existing standards for data monitoring.

To standardize such dashboards, the Supporting Pediatric Research on Outcomes and Utilization of Telehealth (SPROUT)–CTSA Network formed the Telehealth Equity Workgroup. Evidence on best practices for the collection and use of equity-related data continues to evolve. Based on the review of the existing literature and the operational experience of clinicians, informaticists, and researchers in this workgroup, we aim to describe a strategic framework for adult- and pediatrics-serving health systems to execute telehealth equity dashboards through 3 phases: define, design, and deploy (Figure 1). In addition, we offer a checklist for framework navigation (Figure 2) to motivate more critical monitoring and evaluation of health systems' current telehealth practices and ultimately identify service delivery gaps.

**Figure 1.** Telehealth equity dashboard framework.

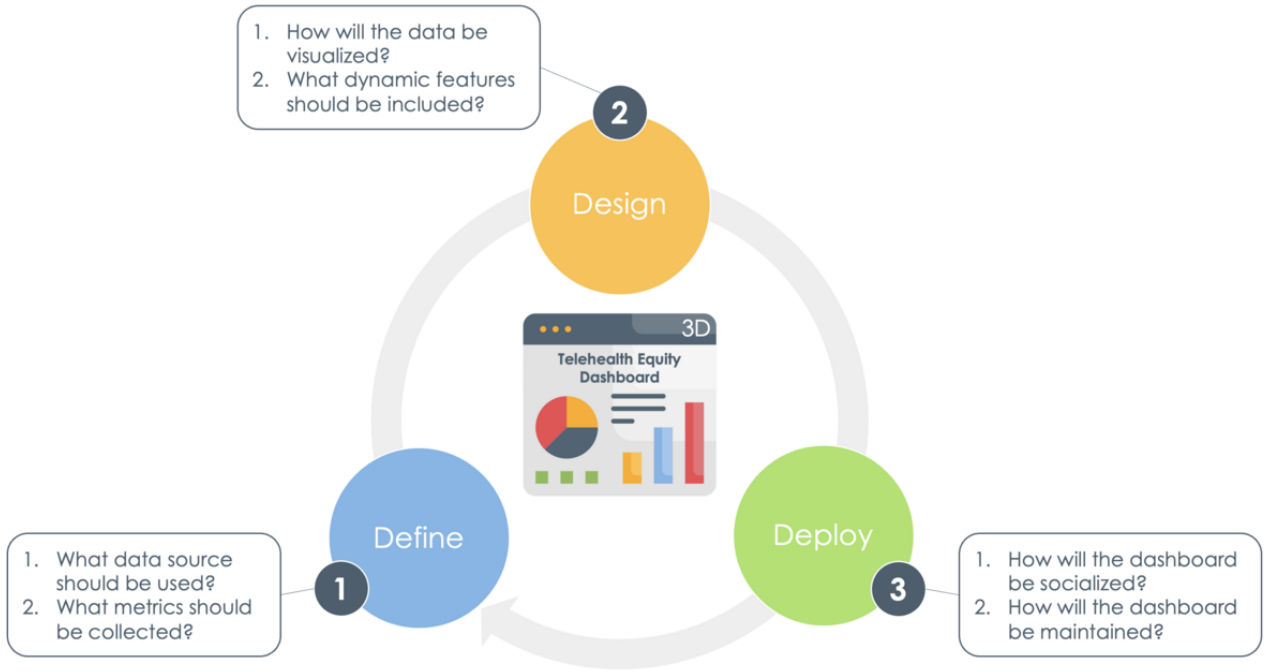


**Figure 2.** Telehealth Equity Dashboard Checklist (aSTEM: SPROUT Telehealth Evaluation and Measurement).

**Phase I: Define**

**Step 1: Identify data sources and collect accurate data on equity stratifiers**

Internal sources:
- ☐ Electronic medical record
- ☐ Patient experience surveys

External sources:
- ☐ National census databases
- ☐ National and regional community surveys
- ☐ News and industry papers

**Step 2: Choose metrics based on STEM[a] framework**

Health outcomes:
- ☐ Individual or population-level disease specific measures (eg, $HbA_{1c}$, BMI percentile, depression screening results, Vanderbilt score, and mortality)

Health delivery:
- ☐ Access (eg, interpreter use and portal use)
- ☐ Effectiveness (eg, guideline adherence)
- ☐ Quality (eg, diagnostic accuracy and % harm)
- ☐ Cost (eg, health care use)

Individual experience:
- ☐ Satisfaction (eg, promoter score)
- ☐ Usability (eg, technical quality)
- ☐ Burden reduction (eg, saved workdays)
- ☐ Adaptability

Key performance indicators:
- ☐ No-show rates
- ☐ Wait times
- ☐ Patient volume
- ☐ Percent telehealth visits

Equity stratifiers:
- ☐ Common demographics (eg, age, gender identity, sex, race, ethnicity, zip code, and census tract)
- ☐ Advanced demographics (eg, language preference, insurance, disability status, and complexity)
- ☐ Technology access (eg, device access, device type, stable internet, and technology literacy)
- ☐ Social determinants of health (eg, economic stability, and neighborhood metrics of disadvantage)

**Phase II: Design**

**Step 1: Create dynamic features**

Recommended features:
- ☐ Comparison groups
- ☐ Variable filter and aggregation functionality
- ☐ Hover functionality to see counts, percents, numerators, and denominators
- ☐ Table displays for detailed views
- ☐ Graphical displays for trends
- ☐ Time range filtering
- ☐ Data missingness reporting

**Step 2: Select visualization tools**

Visualization tools:
- ☐ Epic
- ☐ HealtheIntent
- ☐ Power BI
- ☐ QlikView
- ☐ Qlik Sense
- ☐ R
- ☐ Tableau
- ☐ Other

**Phase III: Deploy**

**Step 1: Socialize dashboard**

Recommended steps:
- ☐ Share with interested parties (leadership, administrative, operations, clinical, patients)
- ☐ Integrate feedback

**Step 2: Maintain data**

Recommended steps:
- ☐ Establish cadence for dashboard tracking and review with each user group
- ☐ Establish cadence for data updates, reporting, and reassessment of performance metrics/goals

## Engaging Interested Parties

Before beginning to create a telehealth equity dashboard, health systems must identify all interested parties to balance diverse perspectives and priorities. This should include all potential dashboard users such as clinical staff, investigators, and administrators as well as dashboard experts and patient advocates. Early engagement facilitates institutional buy-in to both the development and use of a dashboard. In addition, as there is notable variation in data privacy regulations based on patient age, type of medical problem, local health system policy, and federal laws, early involvement of senior leadership can help ensure dashboards are implemented appropriately. Once identified, interested parties must be continuously engaged throughout all phases of the framework process to ensure these dashboards are developed with the intended users in mind.

## Phase 1: Define

First, health systems should consider what data sources to leverage. Data source mapping is one useful technique to identify usable sources for dashboard development. This inventory process involves cataloging all available sources and describing potentially relevant data to allow teams to consider the feasibility, reliability, and quality of these sources [9].

Poor data quality can have negative downstream impacts, as inaccurate or incomplete data can mask disparities [10]. First, patient and caregiver demographics can often be conflated in pediatric and elderly care settings. In addition, previous research found that non-White patients were less likely to have the correct race in their health records and were often mislabeled as White, skewing disparities [11].

Several strategies can mitigate the limitations of missing or inaccurate data [12]. Imputation or Bayesian modeling techniques can help bolster existing data by addressing missingness with inferred values. For example, imputing race and ethnicity identified greater disparities in the COVID-19 pandemic compared with only excluding missing data [13]. Health systems can also enhance existing data by linking their databases to external sources to conduct area-based monitoring [14]. To illustrate, health systems could integrate regional-level population data from national datasets (eg, the National Survey of Children's Health or the American Community Survey for United States health systems) with internal patient data by census tract. Inequities can then be tracked between geographic regions to further support patients from medically underserved areas.

Unfortunately, these methods fail to address the root of data inaccuracy. Improvement of data collection processes is the best long-term solution. Staff training, patient education, and options for self-reporting outside of clinical encounters are the key to improved collection [10]. Greater transparency regarding the purpose of data collection and improved framing of questions to reduce discomfort in sharing sensitive data could also increase self-reporting [11].

Once data sources are established, health systems can select metrics from the domains of the SPROUT Telehealth Evaluation and Measurement Framework [8], including health outcomes (ie, disease-specific measures), health delivery (ie, quality and cost), individual experience (ie, patient experience data), and key performance indicators (ie, implementation measures), as well as equity stratifiers (ie, environmental and patient attributes). In addition, defining each metric's performance target is critical. Targets can be based on peer organizations' performance, past institutional achievements, national-, state-, or county-wide standards, and public policy goals.

## Phase 2: Design

Next, health systems should carefully consider the design of their dashboards, as literature demonstrates how data aggregation and visualization influence the ability to detect disparities. Common broad racial or ethnic categories such as Black or Hispanic obscure within-group differences that can have significant clinical implications [15]. For example, when Asian is grouped with Native Hawaiian and Other Pacific Islanders, such aggregated statistics conceal meaningful differences between subpopulations [16]. Thus, it is important to present data as disaggregated by equity stratifiers as possible, acknowledging that some level of aggregation is necessary given data quality limitations. A recent proposal for revised federal government standards for race or ethnicity classification may guide new best practices [17].

We recommend, at a minimum, comparing data from medically underserved populations tailored to each health system with an aggregated "catch-all" category. Health systems may consider including a reference, which is often the total population, or the group with the largest population, the most favorable health outcomes, or the greatest socioeconomic advantage [18]. However, there are risks of identifying a "reference" group. Selecting White, for example, as the "reference" population may inherently imply "nonreference" populations require assimilation or acculturation or are generally "abnormal."

In addition, designing dashboards with filter functionality across multiple metrics can provide more robust analytics and displays. Irrespective of the population that a health system serves, intersectionality, or the connection between personal identities, is another key attribute to dashboard design, allowing for a more in-depth look at identified disparities. Race as a stratifier on its own could be a proxy for other variables underlying why these disparities exist. However, through filter functionality, users might consider assessing telehealth equity across races with another key attribute such as social determinants of health or internet access [18].

Designers should follow best practices for data visualization [19], including maximizing data-ink ratios and selecting the appropriate software for desired displays. Commercial visualization tools can be found in Figure 2. When choosing visualizations, it is essential to consider ease of interpretation and potential risks of misrepresentation. Tables explicitly lay out comprehensive information but can be difficult to digest. Interpretation can be supported through bolding or color-coding. Graphs can simplify data presentation and draw attention to specific insights, but this simplicity can be misleading [18]. It is essential to include missing data percentages to illustrate uncertainty and incorporate features to understand the context of the data for accurate interpretation. For instance, when interpreting a narrowed disparity, the availability of hover functionality to display numerators, denominators, and count breakdowns for each data point can help users understand the source of this change. In addition to reporting current statistics, the ability to view metrics over time permits the detection of trends and postintervention changes in disparities, which is an essential dashboard function.

Once a preliminary design has been determined, teams can develop a draft dashboard. From this point forward, design and development should proceed concurrently. The draft dashboard should undergo pretesting with sample end users, which can subsequently inform alterations to the design. Keep in mind, multiple designs are likely needed to accommodate different audiences, from frontline staff implementing care and

monitoring day-to-day activity to administrators interested in quarterly or annual trends.

## Phase 3: Deploy

Finally, intentional deployment of a telehealth equity dashboard is critical to increase use, inform and monitor operational and clinical interventions, preserve institutional buy-in, and create a data-driven culture to improve health equity.

Socialization, the process of organizations adjusting to, learning about, and buying into a new initiative, is a key aspect of successful dashboard deployment. Socializing with leadership and clinical providers allows teams to create relationships for long-term reporting and inspires clinicians to use the dashboard in day-to-day operations. Normalizing the use of equity dashboards at all levels can stimulate maintained awareness and action to improve telehealth equity hence laying the foundation for a culture of accountability and quality data collection to address disparities in telehealth and beyond.

In this phase, it is also essential to identify a cadence of dashboard review and updates, given the likely differing preferences among users. For example, leadership may expect a quarterly update on high-level telehealth equity experience, while interpreter services may desire monthly check-ins to monitor progress on their practice changes. Socialization with regular review allows for opportunities for feedback, which studies have shown improve data quality [20]. By recognizing the appropriate set of interested parties, health systems can continue to enhance their dashboards with the right feedback from a broader and inclusive user group.

Once the dashboard has been deployed, data can be used and updated to advocate for new programs or workflows supporting medically underserved populations. The implementation of a dashboard is an ongoing, iterative process through each phase. For example, the telehealth equity dashboard may highlight a disparity that motivates the creation of a new intervention. The implementation of a new intervention may then require new metrics to be added to the existing dashboard or identify other ways to track performance. The dashboard development team may thus return to phase 1 to re-evaluate their sources and metrics. In addition, periodic usability testing by end users can allow for the identification of these key areas of improvement for subsequent iterations. This process, akin to the plan-do-study-act cycle in improvement science, can ensure the adaptability and continual advancement of a dashboard to meet the demands of a dynamic health system [21].

## Call to Action

Dashboards offer an avenue to improve data transparency. Data sharing, especially as it relates to equity, may be limited due to lack of incentives, fear of public scrutiny, or perceived opportunity costs if data are used for research by external parties [22]. However, this creates silos between and even within health systems. Data sharing has the potential to establish shared standards and cross-institutional efforts to improve health on the population level. Therefore, as technology use in health care advances, we must pay close attention to what the data are telling us, be transparent with our progress and shortcomings, and push for change in our care models to ensure equitable quality of and access to care for all patients.

## Conclusions

The COVID-19 pandemic laid bare the implications of the digital divide on health disparities. Nevertheless, telehealth continues to serve as a potential cost-effective care model and promising access point for patients with barriers to in-person services. As such, our strategic framework for developing a telehealth equity dashboard offers a valuable means to track patterns of use and outcomes to provide the evidence needed to support continued investment in an equitable telehealth offering. Telehealth equity dashboards present a promising means to build a culture of data transparency, equity-centered implementation, and continuous improvement to narrow the digital divide and improve access to care for all patients in this expanding world of digital health care.

### Acknowledgments

MJL's contribution to this manuscript was in part funded by the National Institute of Child Health and Human Development Grant (#T32HD060550). JC, Christina Coleman, and the Supporting Pediatric Research on Outcomes and Utilization of Telehealth (SPROUT)–Clinical and Translational Science Awards (CTSA) Network Health Equity Topic Working Group Dashboard's task force's effort contributing to this manuscript was in part funded by National Institutes of Health (NIH)/National Center for Advancing Translational Sciences (NCATS) SPROUT-CTSA Collaborative Telehealth Network Grant (#U01TR002626). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

### Authors' Contributions

MJL contributed to conceptualization, writing of the original draft, visualization, and reviewing and editing. SC managed conceptualization and reviewing and editing. SPG handled conceptualization and reviewing and editing. SMW contributed to conceptualization and reviewing and editing. MA managed conceptualization, writing of the original draft, investigation, and reviewing and editing. JZ contributed to conceptualization and reviewing and editing. JC handled reviewing and editing. PVS managed conceptualization, supervision, project administration, and reviewing and editing.

## Conflicts of Interest

None declared.

## References


1. Roy J, Levy DR, Senathirajah Y. Defining telehealth for research, implementation, and equity. J Med Internet Res 2022;24(4):e35037 [FREE Full text] [doi: 10.2196/35037] [Medline: 35416778]
2. Curfman A, McSwain SD, Chuo J, Yeager-McSwain B, Schinasi DA, Marcin J, et al. Pediatric telehealth in the COVID-19 pandemic era and beyond. Pediatrics 2021;148(3):e2020047795 [FREE Full text] [doi: 10.1542/peds.2020-047795] [Medline: 34215677]
3. Drake C, Lian T, Cameron B, Medynskaya K, Bosworth HB, Shah K. Understanding telemedicine's "New Normal": variations in telemedicine use by specialty line and patient demographics. Telemed J E Health 2022;28(1):51-59 [FREE Full text] [doi: 10.1089/tmj.2021.0041] [Medline: 33769092]
4. Patel SP, Sun E, Reinhardt A, Geevarghese S, He S, Gazmararian JA. Social determinants of digital health adoption: pilot cross-sectional survey. JMIR Form Res 2022;6(12):e39647 [FREE Full text] [doi: 10.2196/39647] [Medline: 36472905]
5. Vaidhyam SAK, Huang K. Social determinants of health and patients' technology acceptance of telehealth during the COVID-19 pandemic: pilot survey. JMIR Hum Factors 2023;10:e47982 [FREE Full text] [doi: 10.2196/47982] [Medline: 37934556]
6. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
7. Gustavson AM, Lewinski AA, Fitzsimmons-Craft EE, Coronado GD, Linke SE, O'Malley DM, et al. Strategies to bridge equitable implementation of telehealth. Interact J Med Res 2023;12:e40358 [FREE Full text] [doi: 10.2196/40358] [Medline: 37184909]
8. Chuo J, Macy ML, Lorch SA. Strategies for evaluating telehealth. Pediatrics 2020;146(5):e20201781 [FREE Full text] [doi: 10.1542/peds.2020-1781] [Medline: 32817398]
9. Hosseinpoor AR, Bergen N, Schlotheuber A. Promoting health equity: WHO health inequality monitoring at global and national levels. Glob Health Action 2015;8:29034 [FREE Full text] [doi: 10.3402/gha.v8.29034] [Medline: 26387506]
10. Wilkins CH, Friedman EC, Churchwell AL, Slayton JM, Jones P, Pulley JM, et al. A systems approach to addressing Covid-19 health inequities. NEJM Catalyst 2021;2(1). [doi: 10.1056/cat.20.0374]
11. Blagev DP, Barton N, Grissom CK, McKee KE, Harrison AM. On the journey toward health equity: data, culture change, and the first step. NEJM Catalyst 2021;2(7). [doi: 10.1056/cat.21.0118]
12. Bilheimer LT, Klein RJ. Data and measurement issues in the analysis of health disparities. Health Serv Res 2010;45(5 Pt 2):1489-1507 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01143.x] [Medline: 21054368]
13. Labgold K, Hamid S, Shah S, Gandhi NR, Chamberlain A, Khan F, et al. Measuring the missing: greater racial and ethnic disparities in COVID-19 burden after accounting for missing race/ethnicity data. medRxiv. Preprint posted online on October 20, 2020 2020 [FREE Full text] [doi: 10.1101/2020.09.30.20203315] [Medline: 33024980]
14. Hosseinpoor AR, Bergen N. Area-based units of analysis for strengthening health inequality monitoring. Bull World Health Organ 2016;94(11):856-858 [FREE Full text] [doi: 10.2471/BLT.15.165266] [Medline: 27821889]
15. Kauh TJ, Read JG, Scheitler AJ. The critical role of racial/ethnic data disaggregation for health equity. Popul Res Policy Rev 2021;40(1):1-7 [FREE Full text] [doi: 10.1007/s11113-020-09631-6] [Medline: 33437108]
16. Shimkhada R, Scheitler AJ, Ponce NA. Capturing racial/ethnic diversity in population-based surveys: data disaggregation of health data for Asian American, Native Hawaiian, and Pacific Islanders (AANHPIs). Popul Res Policy Rev 2021;40(1):81-102. [doi: 10.1007/s11113-020-09634-3]
17. Revesz R. Initial Proposals For Updating OMB's Race and Ethnicity Statistical Standards.: The Federal Register; 2023. URL: https://www.federalregister.gov/documents/2023/01/27/2023-01635/initial-proposals-for-updating-ombs-race-and-ethnicity-statistical-standards [accessed 2024-08-14]
18. Penman-Aguilar A, Talih M, Moonesinghe R, Huang D. Response to scanlan concerning: measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. J Public Health Manag Pract 2016;22(6):614-615. [doi: 10.1097/PHH.0000000000000494] [Medline: 27682733]
19. Midway S. Principles of effective data visualization. Patterns (N Y) 2020;1(9):100141 [FREE Full text] [doi: 10.1016/j.patter.2020.100141] [Medline: 33336199]
20. Taggart J, Liaw S, Yu H. Structured data quality reports to improve EHR data quality. Int J Med Inform 2015;84(12):1094-1098. [doi: 10.1016/j.ijmedinf.2015.09.008] [Medline: 26480872]
21. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ Qual Saf 2014;23(4):290-298 [FREE Full text] [doi: 10.1136/bmjqs-2013-001862] [Medline: 24025320]
22. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. BMC Public Health 2014;14:1144 [FREE Full text] [doi: 10.1186/1471-2458-14-1144] [Medline: 25377061]

## Abbreviations

**CTSA:** Clinical and Translational Science Awards
**DHEF:** Digital Health Equity Framework
**NCATS:** National Center for Advancing Translational Sciences
**NIH:** National Institutes of Health
**SPROUT:** Supporting Pediatric Research on Outcomes and Utilization of Telehealth
**STEM:** SPROUT Telehealth Evaluation and Measurement

*Edited by T de Azevedo Cardoso; submitted 27.02.24; peer-reviewed by R Bidkar, D Chong, J Jansen; comments to author 23.04.24; revised version received 14.05.24; accepted 17.07.24; published 04.09.24.*

*Please cite as:*
*Luke MJ, Craig S, Pak-Gorstein S, Arellano M, Zhang J, Wright SM, Chuo J, Scribano PV*
*Narrowing the Digital Divide: Framework for Creating Telehealth Equity Dashboards*
*Interact J Med Res 2024;13:e57435*
*URL: https://www.i-jmr.org/2024/1/e57435*
*doi:10.2196/57435*
*PMID:*

Viewpoint

# Automated Psychotherapy in a Spaceflight Environment: Advantages, Drawbacks, and Unknowns

Logan Smith[1], PhD

Oklahoma State University, Stillwater, OK, United States

**Corresponding Author:**
Logan Smith, PhD
Oklahoma State University
306 Psychology Building
Stillwater, OK, 74078
United States
Phone: 1 772 242 5012
Email: logan.smith12@okstate.edu

**Related Articles:**

Comment in: https://mhealth.jmir.org/2024/1/e57978/

This is a corrected version. See correction statement: https://www.i-jmr.org/2024/1/e67671

## Abstract

Various behavioral and mental health issues have been reported by space crews for decades, with the overall number of mental health complications expected to be higher than is publicly known. The broad range of mental health complications encountered in space is expected to grow as people venture deeper into space. Issues with privacy, dual relationships, and delayed communications make rendering effective psychological therapy difficult in a spaceflight environment and nearly impossible in deep space. Automated psychotherapy offers a way to provide psychotherapy to astronauts both in deep space and low Earth orbit. Although automated psychotherapy is growing in popularity on Earth, little is known about its efficacy in space. This viewpoint serves to highlight the knowns and unknowns regarding this treatment modality for future deep space missions, and places an emphasis on the need for further research into the applicability and practicality of automated psychotherapy for the spaceflight environment, especially as it relates to long-duration, deep space missions.

***(Interact J Med Res 2024;13:e58803)*** doi:10.2196/58803

**KEYWORDS**

mental health; deep space; astronauts; aerospace medicine; spaceflight; flight; psychotherapy; privacy; communication

## Introduction

The assumed importance of mental health in a spaceflight environment has changed over time, with earlier mission planners thinking that mental health complications would not be an issue for sufficiently qualified personnel [1]. This "right stuff" mentality was thought to indicate that astronaut candidates who were resilient to stress, capable of operating under extreme pressures, and able to draw on a reserve of inner strength to push through dangerous and anxiety-provoking circumstances would be the ideal astronaut in space [2]. However, these highly specific selection criteria do not appear to be capable of picking individuals who are impervious to mental health complications during spaceflight, with a number of notable examples being reported over the past 60 years.

In 1982, a Russian cosmonaut reported struggles with isolation and depressed mood while in space [3]. A 1985 Russian Soyuz mission ended 60% earlier than expected due to what is believed to be issues involving "mood and performance" for crew members [3]. In addition, over the course of 89 American space shuttle missions, there were 34 separate behavioral issues reported among the 208 different crew members [4]. The differing labels used for these instances—hostile and uncooperative, depressed mood, mood and performance, and behavioral issues—reflect the imprecise way in which mental health complications have been discussed in relation to astronaut activities. It is important to note that these instances do not

necessarily reflect something that would have threatened the safety of crew or the mission and were often transient.

Of the psychological and behavioral problems that have been reported and made public, recorded issues include anxiety, depression, irritability, sleep-wake disorders, asthenia, interpersonal tension, impaired judgement, inappropriate behavior, stress, exhaustion, euphoria, neurosis, accentuation of negative personality traits, and various cognitive impairments [4-7]. While these reported psychological and behavioral problems are broad, it should be noted that these have all been reported in missions to low Earth orbit; it is not yet determined what psychological issues could present during long-duration, deep space missions. A general consensus is that all of these previously reported problems could occur in deep space, in addition to serious psychiatric symptoms, psychotic disorders, delirium, and homesickness, among others [4,8-10]. Although these examples are noteworthy, they do not give a complete picture of the mental health incidences that have occurred during spaceflight. It is believed that the actual incidence rate is higher than previously reported and that a number of astronauts may be unwilling to divulge these details with ground personnel [11]. The purpose of this viewpoint is to summarize the current state of automated psychotherapy for the spaceflight environment, especially as it relates to the advantages that it offers, the drawbacks associated with this treatment modality, and the current unknowns.

## Current Treatments for Mental Health Problems in Space

Mental health interventions in a spaceflight environment begin at the crew selection phase, where astronauts are screened for commonly known psychological risk factors [2], commonly referred to as “select-out” logic. In addition, in accordance with “select-in” logic, previous research has indicated that the ideal psychological profile for a crew member for a space mission, especially a long-duration, deep space mission, is an individual with high levels of adaptability, resistance to stress, psychological stability, and strong social skills [12]. It is believed that these highly specific psychological selection criteria may help reduce the rate of psychological issues experienced during the mission. In addition, the Human Behavior and Performance Support Program, a component of NASA comprised of highly trained psychologists, physicians, and other mental health professionals, is tasked with providing support to an astronaut at all phases of the mission—before, during, and afterward. The Human Behavior and Performance Support Program offers entertainment and activities to counteract the detrimental effects of boredom and isolation, regular private video conferences with psychologists on the ground, interactions with family and friends, and other supportive measures to prevent mental health complications from escalating to a degree that may present problems for mission success and could require further treatment.

When a mental health complication is reported during a spaceflight mission, ground personnel, medical personnel, and astronauts have a variety of methods available to them to treat the condition. The first step often focuses on reducing the levels of stress that the astronaut is feeling, as the very environment of a spaceflight mission may produce physical and psychic strain on the astronaut through a combination of microgravity, noise, isolation, reduced privacy, radiation, work-leisure balance, and other factors [1,7,10,11,13,14]. While stress reduction is commonly used as a preventative measure, it is also helpful after symptoms of psychopathology have been reported [8]. One research team has categorized the general stress reduction strategies for a spaceflight environment to show their focus on 4 main areas: ergonomic (factors associated with the design of the spacecraft), physiological (factors associated with nutrition, sleep, and hormonal balance), psychological (factors associated with known psychological stressors), and psychosocial (factors associated with relationships between the astronaut and other crew members as well as ground personnel) [14]. Recommended treatments for dealing with stress while in space often involve increased leisure time, increased time spent communicating with loved ones on Earth, and structured interventions. One such structured intervention is the Spaceflight-Induced Stress Management Plan, which involves a series of group training modules that prepare astronauts for stressors and equips them with the skills needed to form support groups while in space [14]. Given the strong connection between reduced coping with stress and subsequent psychopathology [15], it is evident that the ability to de-escalate the stress that an astronaut is feeling may reduce their likelihood of developing mental health complications, symptoms, and disorders while in space. However, it is important to note that while stress reduction initiatives can help improve the overall mental health of an astronaut, they are not likely to be a sufficient intervention to treat psychopathological responses, such as depression, psychosis, or severe anxiety disorders.

In the event that a mental health complication has been reported that cannot be remedied by noninvasive stress interventions alone, it is recommended that some sort of psychological therapy begin [6]. Regular scheduled interactions between astronauts and a psychologist on the ground may provide a normalized way for crew members to discuss their emotions, relieve tension, and receive feedback from trained specialists [16]. However, the usual version of psychological therapy used on Earth, where a provider and a client have roughly 1 hour of uninterrupted, private, structured discussions using a standardized treatment, is exceedingly difficult in space, and there are little available data on the practice or outcomes of psychological therapy in a spaceflight environment. This will be discussed in further detail in the *Difficulties of Therapy in Space* section. In addition, the use of monitoring tools that can be used to capture intraindividual variability across various behavioral and mental health markers, including real-time physiological measurements and medical monitoring, should be explored as a way to passively and actively gather data on the current functioning of astronauts as a way of detecting potential mental health concerns as early as possible. Often, these tools are used in psychological research to produce reliable data and can involve nonintrusive wearable devices to detect heart rate and skin conductance (such as a ring or a wristband), as well as passive environmental monitors for eye tracking and movement.

A commonly used treatment for mental health complications in space is medication, which is a separate treatment modality from psychotherapy and may be used independent of psychotherapy or in conjuncture with psychotherapy. The current psychiatric formulary onboard the International Space Station includes antidepressants, antipsychotics, anxiolytics, anticholinergics, sleep agents, and wake agents [4]. It is important to note that these medications are helpful for treating the expected psychological disorders that may be encountered in low Earth orbit and do not include medications for more severe psychopathologies, such as more powerful antipsychotics. Rather, this formulary was created with the assumption that severe psychological emergencies may be treated by a rapid return to Earth. Future deep space missions may need a more comprehensive formulary for a setting where a return to Earth is not feasible [4].

Taken together, it is apparent that most interventions for mental health complications in a spaceflight environment have focused on crew selection criteria and methods for coping with stress. While psychiatric medications are available in space and have been used by at least 24 crew members spread out among 20 different missions [17], the efficacy of the gold standard of psychological care, structured psychological therapy, is largely unknown for the spaceflight environment. The following section will focus on this topic.

## Difficulties of Therapy in Space

It is recommended that physicians aboard the International Space Station be trained in psychological therapy in order to provide direct psychological care to crew members with a psychological disorder or symptom [18]. However, the American Psychological Association's Ethical Principles of Psychologists and Code of Conducts outline that a psychologist providing therapy to an individual, while at the same time already possessing another relationship with that person, is engaged in multiple relationships—a potentially dangerous state for a therapist and client to be in that represents real hazards to both individuals [19]. While dual relationships offer a potential pathway for harm to come to either the therapist or the client, they cannot always be avoided [20]. When one considers the perils associated with strained relationships in a spaceflight environment, including lack of cooperation between parties, interpersonal conflicts, reduced privacy and escape, and other hazards, it is apparent that any situation that could produce these hazards should be avoided wherever possible. For these reasons, it may be advisable to avoid providing psychological therapy to a crew member from another crew member on the same mission, if possible.

An alternative to intracrew psychological therapy may be ground-based psychological teletherapy. The proliferation of teletherapy during the COVID-19 pandemic has provided a way for more people to receive psychological care than was previously possible and has acted as an accelerator for this sort of technology [21], acting as a proving ground for the efficacy of teletherapy. However, the application of teletherapy to a spaceflight environment has not been fully explored. Astronauts in space commonly complain about a lack of privacy while onboard a spacecraft or space station [22,23], which may make teletherapy exceedingly difficult while in space. The lack of privacy during therapy, often referred to as a lack of a "safe therapeutic space," can seriously threaten the effectiveness of psychological therapy [24], as a therapy client may be unwilling or unable to divulge crucial details, express emotions fully, or fully participate in therapy in other crucial ways.

Although a lack of privacy in a spaceflight environment can conceivably be corrected through modified living quarters, one issue that cannot be corrected is delayed communications. Previous space simulation studies have revealed that delayed communication creates a host of problems for crews, including confusion and wasted crew time, decreased verbal encoding efficacy, increased stress and frustration, and general task and communication errors [25-28]. Surprisingly, difficulties were seen across the range of possible delayed voice communication times, with issues being found when messages were delayed by as brief as just fractions of a second [26] or as long as 5 minutes [25]. Currently, crews aboard the International Space Station experience communication delays between the crew and the mission support team on the ground measured in the millisecond range, although it is known that this will increase to upwards of 22 minutes for crewed missions to Mars [29], and even longer for more distant destinations. It is apparent that humans are sensitive to voice communications being delayed by any noticeable amount, with difficulties, errors, and stress quickly appearing under these conditions; during an emotionally charged psychotherapy session, these effects may be more pronounced.

Although psychological therapy is considered to be one of the most desirable psychological interventions and is empirically supported for both short-term and long-term outcomes, its use in a spaceflight environment may be difficult. Dual relationships between crew members create barriers to intercrew psychotherapy by introducing increased risk for other negative outcomes. In addition, a combination of a lack of privacy and delayed communications creates significant barriers for teletherapy between ground personnel and astronauts in low Earth orbit. As such, there are limited options available for psychotherapy in low Earth orbit.

## Particular Issues Posed by a Deep Space Environment

A deep space mission may have particular stressors that create an increased risk for mental health complications; in addition to the known stressors of spaceflight (eg, monotony, microgravity, awareness of danger, interpersonal tension, and radiation), a deep space mission may encounter additional psychological stressors, including increased isolation, the psychological effects of distance from Earth, the knowledge of a lack of rescue, prolonged homesickness, and other related phenomena [10,30]. In addition, the very nature of deep space creates a certainty that delayed communications will become a facet of life for deep space astronauts; as the speed of light dictates the maximum speed at which information can travel, messages from Earth to Mars can take up to 22 minutes in 1 direction [29], with longer delays being a certainty for deeper space missions to asteroids, the moons of the gas giants, and

other destinations. As such, delayed communication is expected to be one of the most pressing issues with regard to future long-duration, deep space missions [25]. When combined with the known and expected stressors associated with deep space missions, it is clear that an option for psychological care is needed that does not rely on Earth-based interventions, while also avoiding the dual relationship hazards of intercrew psychotherapy.

## Need for Automated Psychotherapy for Spaceflight

An increasingly common psychological treatment modality, automated psychotherapy, sometimes called computer-mediated psychotherapy, cybertherapy, or computer-created virtual reality for therapy [31,32], is often seen as a method of delivering psychological therapy (psychotherapy) to various groups of people, especially populations that are underserved or otherwise unreachable through traditional therapeutic routes [33]. Automated psychotherapy generally takes the form of a series of computer-delivered modules that explain applicable psychological terms to the therapy client, provides a degree of psychoeducation necessary to progress through the module, delivers homework tasks and the necessary training required for those tasks, and assesses the symptoms and progress of the client using standardized empirically supported measures. Automated psychotherapy is not to be confused with self-help books, which, although found to be effective for temporary amelioration of depressive symptoms (among other disorders), have not been found to produce lasting effects past 6 months [33]. Rather, automated psychotherapy has been shown to be highly effective in treating various psychopathologies with lasting and reliable results [34].

Automated psychotherapy provides a way for astronauts to receive empirically supported psychological therapy while in space and distant from the Earth. Notably, as will be outlined in greater detail below, automated psychotherapy is able to mitigate the issues that confound traditional therapy or teletherapy in a spaceflight environment. There are research projects currently underway to investigate the efficacy of, and design a treatment for, automated psychotherapy in a spaceflight environment [35]; these findings will be explored in greater detail below. Notably, future deep space missions will likely require an automated psychotherapy option that can function independent of ground-based personnel, although automated psychotherapy options in the foreseeable future will likely involve a trained human therapist to some extent or another. Discussions of automated psychotherapy below are generally in reference to the more automated end of this treatment modality.

## Advantages of Automated Psychotherapy for Spaceflight

Automated psychotherapy has been widely discussed as a potentially popular method of delivering psychotherapy on Earth, as it allows treatments to be tailored to the client in a way that reduces the workload on experienced clinicians [36]. Cognitive behavioral therapy (CBT) is a common psychological treatment modality that is often modified into an automated psychotherapy course, usually referred to as internet CBT (iCBT). Here, it should be noted that the CBT umbrella is broad; many different components of other treatment modalities (eg, mindfulness-based treatments for anxiety disorders and exposure-based treatments for anxiety and trauma disorders) can fit under this umbrella, which is why CBT is often considered to be the gold standard of frontline psychological care [37]. A recent review has found that iCBT has been an effective treatment for a variety of psychopathologies, including depression, generalized anxiety disorder, panic disorder, obsessive compulsive disorder, posttraumatic stress disorder, adjustment disorder, bipolar disorder, and phobias, among others [38]. Recently, research has indicated that mindfulness-based interventions for stress, anxiety, depression, and other psychological disorders may be particularly useful in the spaceflight environment [39]. Given that automated psychotherapy has already been tested for mindfulness training, this may be a particularly useful function of this treatment modality [40]. In future spaceflight environments, automated psychotherapy may be the only form of treatment available to astronauts, especially during deep space missions [14]. Based on the empirical support for the use of automated psychotherapy for a wide range of psychopathologies, it is evident that automated psychotherapy has broad applicability and can be tailored to the individual in a flexible and supportive way [36]. In addition, the instantaneous nature of automated psychotherapy, and the lack of a need for a second party (the trained clinician), means that this treatment modality can easily be integrated into the astronaut's schedule in a way that is convenient and accessible to them, reducing the barriers to treatment. Given that a lack of convenience is a commonly cited factor for not pursuing psychotherapy on Earth [36], reducing this barrier in a spaceflight environment is of the utmost importance.

In addition, automated psychotherapy provides a way for therapy to be conducted that does not require individuals to enter dual relationships (eg, both as professional colleagues and as therapist or client) [41]. In a spaceflight environment, the importance of this cannot be overstated. Given the known history of intracrew social tensions to quickly devolve the general morale and effectiveness of crews in space, any steps that can reduce the risk of social strain on space crews are steps worth taking. Similarly, the increased privacy of automated psychotherapy treatments can also protect the social reputation of crew members and reduce the likelihood of conflict. Given that some forms of automated psychotherapy allow for a client to undergo therapy without having to verbally speak, the opportunity for inadvertent or intentional eavesdropping from other crew members is significantly reduced. Accordingly, automated psychotherapy clients may feel more comfortable divulging their thoughts, feelings, and emotions in a therapeutic context without fear of their reports being overheard by other crew members. Notably, astronauts have previously cited a lack of privacy as a barrier to participating in psychosocial research while onboard the International Space Station [42], and these concerns can be expected to persist in a therapeutic environment.

Automated psychotherapy also creates the only pathway for therapy to occur between the client and another party not on the spacecraft that does not result in delayed communications. Given that an accumulation of stressors can threaten the therapeutic relationship and the overall effectiveness of psychotherapy [43], any steps that may reduce these stressors are worth pursuing. Automated psychotherapy offers a way to reduce the strain caused by delayed communications, protects the privacy of the client, reduces the risk of dual relationships, offers a tailored and customizable course of treatment, and has broad applicability for a wide range of psychopathologies. Accordingly, this treatment modality appears to be useful for future long-duration, deep space missions and may even offer advantages to crew members in low Earth orbit. As such, the usefulness of automated psychotherapy in a spaceflight environment should be further investigated.

## Drawbacks of Automated Psychotherapy for Spaceflight

However, automated psychotherapy is not impervious to all criticisms, and there are noteworthy concerns about this treatment modality. One potential detractor for automated psychotherapy is its apparently low treatment adherence rate compared with other psychotherapies; a meta-analysis of iCBT and face-to-face CBT adherence rates found that 84.7% of individuals receiving face-to-face CBT would complete their treatment, compared with 65.1% of individuals receiving iCBT [44]. Little data have been found regarding generalized predictors of treatment nonadherence for automated psychotherapy [45]. However, incorporating knowledge from the medical field more broadly may provide some insight to treatment nonadherence for automated psychotherapy. For example, the Medication Adherence Model provides a framework for understanding nonadherence to medication recommendations and treatments [46]. Within this model, it is thought that 2 different types of nonadherence may be seen: the intentional decision to miss a medication dose and the unintentional interruptions that can cause a medication dose to be missed. A core concept related to nonadherence is whether the patient shows purposeful actions to increase adherence, can demonstrate patterned behaviors to increase adherence, and are receptive to feedback to increase adherence. The Medication Adherence Model incorporates concepts from various cognitive and self-regulatory models of behavior to explain the processes involved in medication adherence [46]. Considering the framework provided by the Medication Adherence Model, it is possible that adherence to psychological therapies, and therefore, automated psychotherapies, may be impacted by similar factors. With regard to adherence to automated psychotherapy, it is possible that individuals who show purposeful actions to adhere to their treatment, can implement patterned behaviors to help adhere to their treatment, and are receptive to feedback regarding their treatment adherence may be able to complete automated psychotherapy modules at a higher rate than was previously found. The proposed ideal psychological profile for future deep space astronauts includes traits similar to these [12], which may indicate that future deep space astronauts will have a higher adherence rate to automated psychotherapies than the general population. However, an inclination to adhere to therapy provided by an automated psychotherapy system may be moderated by the astronaut's belief that the information received from the system is accurate and helpful; if a client does not believe this to be the case, treatment adherence may be low. Accordingly, it is important to investigate how future deep space astronauts feel about the usefulness of these therapy solutions, and effort should be made to provide education surrounding the use of, and outcomes associated with, automated psychotherapy.

Automated psychotherapy may also be unsuitable for severe psychopathologies. Given that automated psychotherapy requires that an individual be able to self-monitor, adhere to their own treatment, and provide accurate assessments of their own thoughts and feelings, certain types of delirium, psychosis, or other severe psychopathologies may render an individual unable to use automated psychotherapy treatments. However, ruling out a treatment modality based on the severity of a diagnosis is common practice in therapeutic settings [47]. In the case that a crew member is experiencing a severe psychopathology, it is possible automated psychotherapies could be modified to incorporate another crew member to help provide the treatment. This, combined with the psychiatric formulary available to the crew, may offer a way to treat an individual who is far from Earth and is experiencing severe psychological symptoms.

It is also possible that while automated psychotherapy is intended to alleviate some of the uneasiness that an astronaut may feel about divulging sensitive mental health concerns to another person, this treatment modality may be unable to fully address this concern. For example, an astronaut may think that informing an automated psychotherapy program about their thoughts of depression, their anxiety surrounding a mission objective, or their feelings of isolation and loneliness will somehow be stored in an accessible personnel file or be relayed to mission planners on the ground. These concerns may prevent some astronauts from seeking help, and addressing these concerns will take deliberate education and planning to ensure that astronauts understand the full limits of confidentiality between themselves and the automated psychotherapist.

## First Examples of Automated Psychotherapy Options for Spaceflight

Numerous early, simplified examples of automated psychotherapy have been designed and tested for future long-duration, deep space missions. One such project aims to create a digital tool that allows astronauts to monitor their behavior, performance, and feelings; make small changes to their routines as recommended; and reduce stress in key areas to help reduce the risk of developing a psychopathology [48]. Similarly, an interactive media program is being developed to help astronauts cope with interpersonal conflict and depression while engaged in deep space missions [49]. This program was tested by a crew of 6 individuals who spent 8 months in group isolation in a space analog environment [50]. Overall, the space analog crew found the conflict and stress modules of the treatment to be particularly helpful, with one of the most cited areas of improvement being a desire to learn how to use these modules to help other crew members. In addition, there are

various virtual and augmented reality methods for mitigating psychological demands under development and in various stages of testing. One such option is the Crew Interactive Mobile Companion, developed by IBM and Airbus Group and already tested on the International Space Station [51]. Crew Interactive Mobile Companion is designed as an artificial intelligence assistant for astronauts that is designed to offer them guidance on certain tasks and supply answers to technical questions. Overall, the development of automated psychotherapy for a spaceflight environment is still in its early stages. In addition, it should be noted that there is not public-facing information about what, if any, automated psychotherapy tools are currently available to astronauts in space. NASA and other space agencies should make a concerted effort to publish this information if it is available; if no automated psychotherapy tools are currently being used or tested in an actual spaceflight environment, efforts should be made to begin incorporating these as soon as possible.

## Variables Still to be Determined About Automated Psychotherapy for Spaceflight

Although much progress is being made with regard to developing automated psychotherapy treatments for long-duration, deep space missions, there are still key variables that have yet to be fully explored. More information is needed regarding adherence to automated psychotherapy, especially as it relates to the common personality profiles of astronaut crews. While the Medication Adherence Model offers general principles for treatment adherence [46], the applicability of this model to automated psychotherapy for astronauts has yet to be explored. There is also a need to investigate whether the usefulness of automated psychotherapy could expand the astronaut selection criteria. Currently, some of the astronaut selection criteria operate on a "rule-out" procedure, where certain medical, occupational, sociological, and psychological factors are searched for in order to disqualify a particular astronaut candidate [52]. Identifying the cognitive components most associated with treatment adherence to automated psychotherapy courses could help add to the existing "rule-in" procedures for astronaut selection (involving required skills or attributes for a given mission), wherein certain cognitive factors that would indicate that an astronaut candidate is more likely to respond to automated psychotherapy courses in the future are used when creating astronaut crew rosters. Given that it is almost expected that an astronaut will experience some level of psychological distress, discomfort, or disorder during future long-duration, deep space missions [53], perhaps it is worth searching for astronauts who best respond to psychological treatments, rather than restricting astronaut classes to only those who we believe are the most resilient against psychological disorders.

## Fitting Automated Psychotherapy Into Mental Health for Spaceflight More Broadly

Recent publications involving the mental health of the spaceflight environment more broadly have provided an excellent framework for understanding what our current knowledgebase is for space psychology and what research is still needed [41]. Notably, automated digital therapeutics has been recognized as a potentially useful psychological support tool for long-duration, deep space missions [54], indicating that multiple research teams have identified this as a helpful tool that should be developed. As research into digital therapeutics, a line of medical interventions that allow patients to interact with digital health technology in lieu of or in addition to medical professionals, accelerates, it is hoped that these technologies will continue to proliferate [55]. Automated psychotherapy can be conceptualized as an arm of digital therapeutics that has been proven to be particularly effective on Earth [34] and may drastically improve psychological outcomes for future astronauts that experience mental health difficulties.

## Conclusion

Automated psychotherapy offers a way for astronauts engaged in long-duration, deep space missions to receive empirically supported psychotherapy in a way that protects their privacy, reduces the risk of interpersonally straining dual relationships between crew members, and removes the difficulties presented by delayed communications. Although there are little data available regarding predictors of adherence to automated psychotherapy treatments, knowledge incorporated from the Medication Adherence Model suggests that the cognitive factors associated with reduced adherence to medical treatments may be less of a risk for astronauts due to the psychological selection criteria used by space agencies. The usefulness of automated psychotherapies for severe psychopathologies and unwilling participants has yet to be fully explored, but its overall efficacy for a broad range of psychopathologies that could be encountered during a long-duration, deep space mission indicates that automated psychotherapy could be a useful tool for safeguarding the mental health of future astronauts.

### Acknowledgments

The author would like to thank Dr Nick Kanas for providing the inspiration for this review. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

### Conflicts of Interest

None declared.

### References

1. Suedfeld P. Invulnerability, coping, salutogenesis, integration: four phases of space psychology. Aviat Space Environ Med 2005;76(6 Suppl):B61-B66. [Medline: 15943196]
2. Santy PA. Choosing the Right Stuff: The Psychological Selection of Astronauts and Cosmonauts. Westport, CT: Praeger Publishers/Greenwood Publishing Group; 1994.
3. Buckey JC. Space physiology. In: Evidence Report: Risk of Adverse Cognitive or Behavioral Conditions and Psychiatric Disorders. Houston, TX: National Aeronautics and Space Administration; Apr 11, 2006.
4. Friedman E, Bui B. A psychiatric formulary for long-duration spaceflight. Aerosp Med Hum Perform 2017;88(11):1024-1033. [doi: 10.3357/AMHP.4901.2017] [Medline: 29046178]
5. Collins DL. Psychological issues relevant to astronaut selection for long-duration space flight: a review of the literature. Hum Perf Extrem Environ 2003;7(1):43-67 [FREE Full text] [doi: 10.7771/2327-2937.1021] [Medline: 14552348]
6. Kanas N. Psychosocial support for cosmonauts. Aviat Space Environ Med 1991;62(4):353-355. [Medline: 2031641]
7. Morphew E. Psychological and human factors in long duration spaceflight. McGill J Med 2020;6(1):74-80 [FREE Full text]
8. Geuna S, Brunelli F, Perino MA. Stressors, stress and stress consequences during long-duration manned space missions: a descriptive model. Acta Astronaut 1995;36(6):347-356. [doi: 10.1016/0094-5765(95)00115-8] [Medline: 11540600]
9. Oluwafemi FA, Abdelbaki R, Lai JCY, Mora-Almanza JG, Afolayan EM. A review of astronaut mental health in manned missions: potential interventions for cognitive and mental health challenges. Life Sci Space Res (Amst) 2021;28:26-31. [doi: 10.1016/j.lssr.2020.12.002] [Medline: 33612177]
10. Smith LM. The psychology and mental health of the spaceflight environment: a scoping review. Acta Astronaut 2022;201:496-512. [doi: 10.1016/j.actaastro.2022.09.054]
11. Institute of Medicine (US) Committee on Creating a Vision for Space Medicine During Travel Beyond Earth Orbit. In: Ball JR, Evans CH, editors. Safe Passage: Astronaut Care for Exploration Missions. Washington, DC: National Academies Press; 2001.
12. Ursin H, Comet B, Soulez-Larivière C. An attempt to determine the ideal psychological profiles for crews of long term space missions. Adv Space Res 1992;12(1):301-314. [doi: 10.1016/0273-1177(92)90298-c] [Medline: 11536972]
13. Smith L. Space station and spacecraft environmental conditions and human mental health: specific recommendations and guidelines. Life Sci Space Res (Amst) 2024;40:126-134. [doi: 10.1016/j.lssr.2023.10.001] [Medline: 38245337]
14. Monzani L, Kozusznik M, Ripoll P, van Dick R, Peiró JM. Coping in the final frontier: an intervention to reduce spaceflight-induced stress. Psychol 2019;62(1):57-77. [doi: 10.14195/1647-8606_62-1_3]
15. Zimmer-Gembeck MJ, Skinner EA. The development of coping: implications for psychopathology and resilience. Dev Psychol 2016:1-61. [doi: 10.1002/9781119125556.devpsy410]
16. Manzey D, Schiewe A, Fassbender C. Psychological countermeasures for extended manned spaceflights. Acta Astronaut 1995;35(4-5):339-361. [doi: 10.1016/0094-5765(95)98736-s] [Medline: 11541475]
17. Wotring VE. Medication use by U.S. crewmembers on the international space station. FASEB J 2015;29(11):4417-4423. [doi: 10.1096/fj.14-264838] [Medline: 26187345]
18. Kanas N. Psychosocial training for physicians on board the space station. Aviat Space Environ Med 1988;59(5):456-457. [Medline: 3390103]
19. Ethical principles of psychologists and code of conduct. Washington, DC: American Psychological Association; 2017. URL: https://www.apa.org/ethics/code [accessed 2024-09-24]
20. Gottlieb MC. Avoiding exploitive dual relationships: a decision-making model. Psychotherapy (Chic) 1993;30(1):41-48. [doi: 10.1037/0033-3204.30.1.41] [Medline: 11656361]
21. Taylor CB, Fitzsimmons-Craft EE, Graham AK. Digital technology can revolutionize mental health services delivery: the COVID-19 crisis as a catalyst for change. Int J Eat Disord 2020;53(7):1155-1157 [FREE Full text] [doi: 10.1002/eat.23300] [Medline: 32449523]
22. Manzey D, Lorenz B. Human performance during spaceflight. Hum Perf Extrem Environ 1999;4(1):8-13. [Medline: 12182201]
23. Kanas N. Psychological and interpersonal issues in space. Am J Psychiatry 1987;144(6):703-709. [doi: 10.1176/ajp.144.6.703] [Medline: 2884888]
24. Maier CA, Riger D, Morgan-Sowada H. "It's splendid once you grow into it:" client experiences of relational teletherapy in the era of COVID-19. J Marital Fam Ther 2021;47(2):304-319 [FREE Full text] [doi: 10.1111/jmft.12508] [Medline: 33721348]
25. Love SG, Reagan ML. Delayed voice communication. Acta Astronaut 2013;91:89-95. [doi: 10.1016/j.actaastro.2013.05.003]
26. Krauss RM, Bricker PD. Effects of transmission delay and access delay on the efficiency of verbal communication. J Acoust Soc Am 1967;41(2):286-292 [FREE Full text]
27. Kintz NM, Chou C, Vessey WB, Leveton LB, Palinkas LA. Impact of communication delays to and from the international space station on self-reported individual and team behavior and performance: a mixed-methods study. Acta Astronautica 2016;129:193-200. [doi: 10.1016/j.actaastro.2016.09.018]
28. Fischer U, Mosier K, Orasanu J. The impact of transmission delays on mission control-space crew communication. Proc Hum Factors Ergon Soc Annu Meet 2013;57(1):1372-1376 [FREE Full text] [doi: 10.1177/1541931213571303]

29. Byrnes DV, Longuski JM, Aldrin B. Cycler orbit between Earth and Mars. J Spacecr Rockets 1993;30(3):334-336. [doi: 10.2514/3.25519]
30. Kanas N. Psychology in deep space.: The British Psychological Society; 2015. URL: https://www.bps.org.uk/psychologist/psychology-deep-space [accessed 2024-09-24]
31. Binik YM, Cantor J, Ochs E, Meana M. From the couch to the keyboard: psychotherapy in cyberspace. In: Kiesler S, editor. Culture of the Internet. London, England: Psychology Press; 1997:71-102.
32. Suler JR. Psychotherapy in cyberspace: A 5-dimensional model of online and computer-mediated psychotherapy. Cyberpsychol Behav 2000;3(2):151-159. [doi: 10.1089/109493100315996]
33. Parks AC, Szanto RK. Assessing the efficacy and effectiveness of a positive psychology-based self-help book. Ter Psicol 2013;31(1):141-149. [doi: 10.4067/s0718-48082013000100013]
34. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
35. Chang DD, Storch EA, Black L, Berk M, Pellis N, Lavretsky H, et al. Promoting tech transfer between space and global mental health. Aerosp Med Hum Perform 2020;91(9):737-745. [doi: 10.3357/AMHP.5589.2020] [Medline: 32867906]
36. Andrews G, Williams AD. Internet psychotherapy and the future of personalized treatment. Depress Anxiety 2014;31(11):912-915. [doi: 10.1002/da.22302] [Medline: 25407580]
37. David D, Cristea I, Hofmann SG. Why cognitive behavioral therapy is the current gold standard of psychotherapy. Front Psychiatry 2018;9:4 [FREE Full text] [doi: 10.3389/fpsyt.2018.00004] [Medline: 29434552]
38. Kumar V, Sattar Y, Bseiso A, Khan S, Rutkofsky IH. The effectiveness of internet-based cognitive behavioral therapy in treatment of psychiatric disorders. Cureus 2017;9(8):e1626 [FREE Full text] [doi: 10.7759/cureus.1626] [Medline: 29098136]
39. Pagnini F, Thoolen S, Smith N, van Ombergen A, Grosso F, Langer E, et al. Mindfulness disposition as a protective factor against stress in Antarctica: a potential countermeasure for long-duration spaceflight? J Environ Psychol 2024;94:102254. [doi: 10.1016/j.jenvp.2024.102254]
40. Chan S, Li L, Torous J, Gratzer D, Yellowlees PM. Review and implementation of self-help and automated tools in mental health care. Psychiatr Clin North Am 2019;42(4):597-609. [doi: 10.1016/j.psc.2019.07.001] [Medline: 31672210]
41. Trachsel M, Gaab J, Tekin Ş, Biller-Andorno N, Sadler JZ. Why ethics matter in psychotherapy. In: Oxford Handbook of Psychotherapy Ethics. Oxford, United Kingdom: Oxford University Press; 2021:3-30.
42. Ritsher JB, Kanas N, Saylor S. Maintaining privacy during psychosocial research on the international space station. Journal of Human Performance in Extreme Environments 2005;8(1):3. [doi: 10.7771/2327-2937.1041]
43. Pagnini F, Manzey D, Rosnet E, Ferravante D, White O, Smith N. Human behavior and performance in deep space exploration: next challenges and research gaps. NPJ Microgravity 2023;9(1):27 [FREE Full text] [doi: 10.1038/s41526-023-00270-7] [Medline: 36997549]
44. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
45. Chen H, Rodriguez MA, Qian M, Kishimoto T, Lin M, Berger T. Predictors of treatment outcomes and adherence in internet-based cognitive behavioral therapy for social anxiety in China. Behav Cogn Psychother 2020;48(3):291-303. [doi: 10.1017/S1352465819000730] [Medline: 31928568]
46. Johnson MJ. The medication adherence model: a guide for assessing medication taking. Res Theory Nurs Pract 2002;16(3):179-192. [doi: 10.1891/rtnp.16.3.179.53008] [Medline: 12472293]
47. Gilbert SP. Ethical issues in the treatment of severe psychopathology in university and college counseling centers. J Couns Dev 1992;70(6):695-699. [doi: 10.1002/j.1556-6676.1992.tb02148.x] [Medline: 16032825]
48. DRiFT Research and Theory. URL: https://www.driftextremes.com/the-tool [accessed 2022-01-02]
49. Carter JA, Buckey JC, Greenhalgh L, Holland AW, Hegel MT. An interactive media program for managing psychosocial problems on long-duration spaceflights. Aviat Space Environ Med 2005;76(6 Suppl):B213-B223. [Medline: 15943215]
50. Anderson AP, Fellows AM, Binsted KA, Hegel MT, Buckey JC. Autonomous, computer-based behavioral health countermeasure evaluation at HI-SEAS mars analog. Aerosp Med Hum Perform 2016;87(11):912-920. [doi: 10.3357/AMHP.4676.2016] [Medline: 27779949]
51. CIMON (Crew Interactive Mobile Companion). TAdviser. URL: https://tadviser.com/index.php/Product:CIMON_(Crew_Interactive_Mobile_Companion) [accessed 2022-11-03]
52. Barrett JD, Holland AW, Vessey WB. Identifying the "Right Stuff": an exploration-focused astronaut job analysis. 2015 Presented at: 30th annual conference of the Society for Industrial and Organizational Psychology; April 23-25, 2015; Philadelphia, PA URL: https://ntrs.nasa.gov/citations/20140011498
53. Kanas N. Psychological, psychiatric, and interpersonal aspects of long-duration space missions. J Spacecr Rockets 1990;27(5):457-463. [doi: 10.2514/3.26165] [Medline: 11537616]
54. Smith N, Peters D, Jay C, Sandal GM, Barrett EC, Wuebker R. Off-world mental health: considerations for the design of well-being-supportive technologies for deep space exploration. JMIR Form Res 2023;7:e37784 [FREE Full text] [doi: 10.2196/37784] [Medline: 36787162]

55. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023;6(1):38 [FREE Full text] [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]

## Abbreviations

**CBT:** cognitive behavioral therapy
**iCBT:** internet cognitive behavioral therapy

*Edited by T de Azevedo Cardoso; submitted 25.03.24; peer-reviewed by F Pagnini, I Barkaszi; comments to author 05.07.24; revised version received 22.08.24; accepted 29.08.24; published 09.10.24.*

*Please cite as:*
*Smith L*
*Automated Psychotherapy in a Spaceflight Environment: Advantages, Drawbacks, and Unknowns*
*Interact J Med Res 2024;13:e58803*
*URL: https://www.i-jmr.org/2024/1/e58803*
*doi:10.2196/58803*
*PMID:39382952*

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 Silaule[1], MSc; Daleen Casteleijn[2], PhD; Fasloen Adams[3], PhD; Nokuthula Gloria Nkosi[4], PhD

[1]Department of Occupational Therapy, University of the Witwatersrand, Johannesburg, South Africa

[2]Department of Occupational Therapy, University of Pretoria, Pretoria, South Africa

[3]Division of Occupational Therapy, Stellenbosch University, Cape Town, South Africa

[4]Department 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

***(Interact J Med Res 2024;13:e48587)*** doi:10.2196/48587

**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

schizophrenia in rural South Africa and found that this sample has a higher burden of care. A recommendation from this study was to develop strategies to support caregivers with the aim of improving their quality of life.

### 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, social, 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 synonyms 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'] AND ['underserved' or 'deprived' or 'middle-income' or low income country]." On advice from the librarian, gray literature such as conference presentations and unpublished studies was searched on ProQuest Dissertations and Theses Global and public health and conference databases. Finally, 2 reviewers (OS and DC) screened the reference lists of the included articles for additional studies. Textbox 1 presents the final list of search terms used to search published and unpublished studies across all the selected databases in this review. The identified references were imported into the Mendeley reference manager (Elsevier) for screening.

**Textbox 1.** Search terms used in the databases.

'Informal caregiver*' or Caregiver* or Carers or 'Informal Carers' or 'Male Caregiver' or 'Women Caregiver' or 'Family Caregivers' or 'Unpaid Caregiver' or 'Spouse Caregiver' or 'Caretaker' or 'Older Caregiver'

AND

'Caregiver* burden' or 'Caregiver* stress' or '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 'Caretaker* load' or '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 (countr* 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 (countr* OR nation* or population* OR world)) OR ((developing OR (less* N1 developed) OR "under developed" OR underdeveloped

## Source of Evidence Selection

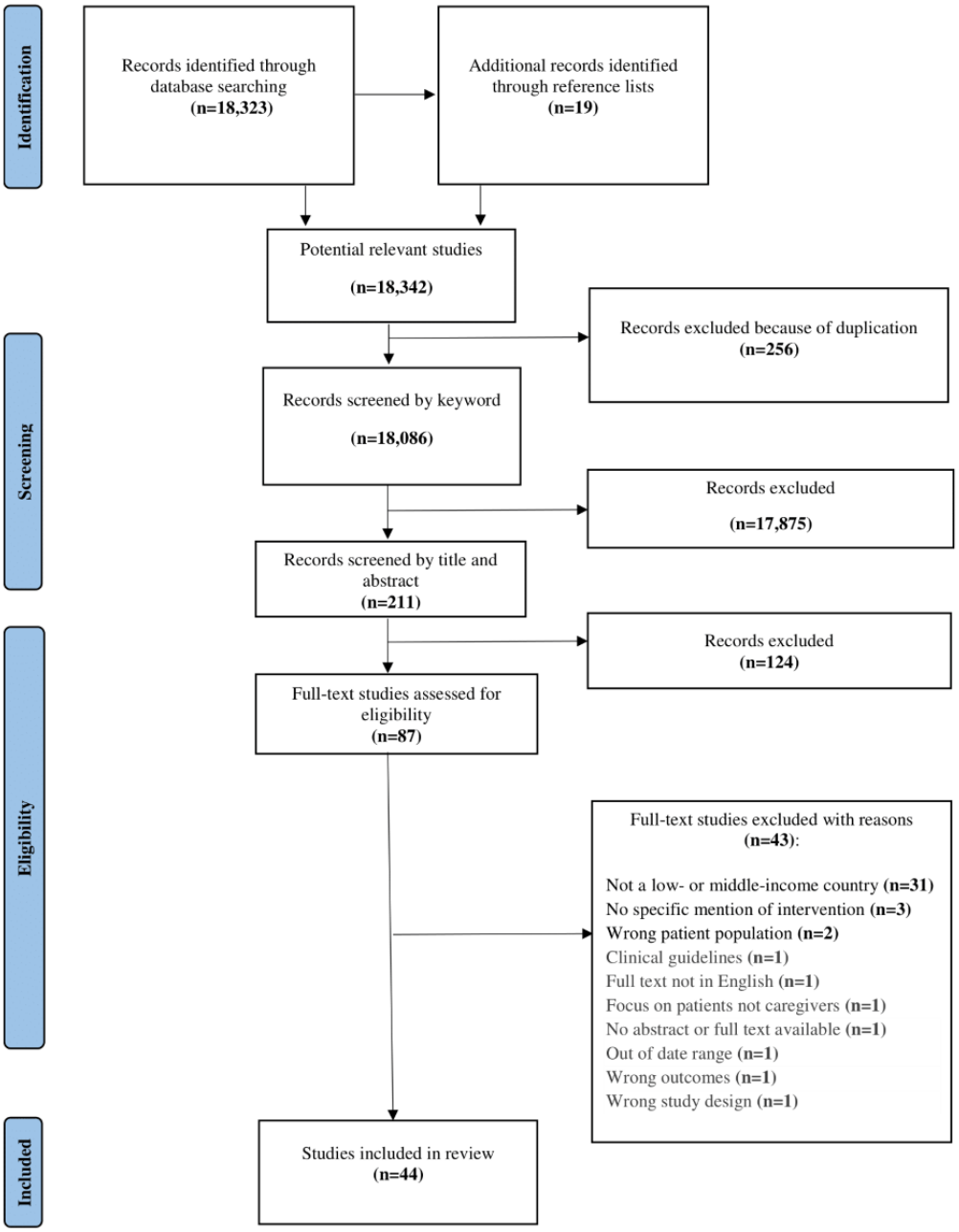
Following the search, all identified citations were exported to the Mendeley reference manager and thereafter to a web-based software, Covidence (Veritas Health Innovation), for primary screening and data extraction from the selected articles. Duplicates were removed, and 2 reviewers (OS and DC) independently 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).

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

### *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).

| Continent | Implemented strategies, n (%) | Recommended strategies, n (%) |
|---|---|---|
| Asia | 24 (83) | 9 (60) |
| Africa | 1 (3) | 3 (20) |
| Europe | 2 (7) | 0 (0) |
| South and North America | 2 (7) | 2 (13) |
| Mixed countries | 0 (0) | 1 (7) |

### *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).

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

## 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 4.** Description of the identified evaluated strategies for alleviating informal caregiver burden.

| Type of strategy | Evaluated interventions | Author-reported outcomes |
|---|---|---|
| Community-based interventions | • CoMHIP[a]<br>• RESHAPE[b] | • Positive outcomes (positive effects on family caregivers) |
| Psychoeducation interventions | • Psychoeducation program<br>• Group psychoeducation<br>• FLEP[c]<br>• Peer-assisted education | • Improved coping skills<br>• Enhanced recovery<br>• Improved QoL[d]<br>• Knowledge and skill acquisition<br>• Reduced burden of care<br>• Reduced anxiety and stress<br>• Reduced psychological strain<br>• Enhanced family functioning<br>• Relapse prevention<br>• Reduced prolonged admissions<br>• Negative outcomes (no efficacy on burden of care, QoL, or self-esteem among BMD[e] caregivers) |
| Support groups | • FPGP[f]<br>• Family-led mutual support program | • Knowledge and skill acquisition<br>• Enhance help seeking<br>• Alleviation of guilt<br>• Improved family and patient functioning<br>• Decreased demand on mental health services<br>• Improved psychosocial health<br>• Improved QoL |
| Cognitive behavioral therapy | • Emotional regulation training | • Reduced burden of care<br>• Reduced anxiety and stress<br>• Increased resilience |
| Spirituality-based intervention | • Spirituality-based program | • Reduced anxiety and stress |
| Guided self-help interventions | • Manual-guided PBSP[g]<br>• The Good Mood Guide | • Strengthened positive caregiving experience<br>• Reduced negative caregiving experience<br>• Increased access to support |
| Smartphone-based interventions | • MHapps[h] | • None |
| Mindfulness and empowerment interventions | • IEP[i]<br>• MBSR[j] program | • Caregiver empowerment<br>• Improved QoL<br>• Decreased depressive symptoms<br>• Increased self-efficacy |
| Collaborative interventions | • Participatory care model | • Reduced burden of care<br>• Increased resilience |
| Standard care | • Multimodal intervention (general medicine, psychiatry, psychology, family therapy, neuropsychological rehabilitation, and occupational therapy) | • Reduced burden of care<br>• Increased social support |

[a]CoMHIP: community mental health early intervention project.

[b]RESHAPE: Reducing Stigma Among Healthcare Providers.

[c]FLEP: Family Link Education Programme.

[d]QoL: quality of life.

[e]BMD: bipolar mood disorder.

[f]FPGP: Family-led Peer Support Group Program.

[g]PBSP: problem-solving–based self-learning program.

[h]MHapps: mental health apps.

[i]IEP: integrated empowerment program.

[j]MBSR: mindfulness-based stress reduction.

**Textbox 2.** Description of the identified recommended strategies for alleviating informal caregiver burden.

**Community-based interventions**

- Strengthening existing primary health care system
- Establishment of day nursing or care
- Respite care
- Vocational training for patients
- Single-family caregiver organization
- Supported employment for patients
- Home visits

**Psychoeducation interventions**

- Group psychoeducation
- Family education
- Continuing education programs for effective practice
- Educational lectures
- Psychoeducation program

**Support groups**

- Nongovernmental mental health–related support groups
- Family caregiver assistance programs

**Cognitive behavioral therapy**

- Family therapy groups

**Spirituality-based intervention**

- Turning to traditional healers and spiritual leaders

**Smartphone-based interventions**

- Mental health apps

**Mindfulness and empowerment interventions**

- Stress management training

**Collaborative interventions**

- Caregiver involvement in all program elements
- Pluralistic and ecological approach to service delivery

**Standard care**

- Telepsychiatry
- Availability of emergency teams

**Financial support**

- Disability grant
- Fee-free mental health services
- Medical insurance and free medication

**Social support**

- Medical care social security
- Social resources

**Counseling**

- 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).

### *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).

| Study | Year of publication | Type of intervention | Type of burden targeted by the intervention | Intervention developer | Intervention deliverer | Number of sessions | Duration of intervention | Frequency of sessions per week | Length of the sessions |
|---|---|---|---|---|---|---|---|---|---|
| Ng et al [56] | 2018 | CoMHIP[a] | Subjective and objective | Not specified | The Integrated Community Centre for Mental Wellness | Not specified | Not specified | Not specified | Not specified |
| Rai et al [55] | 2018 | RESHAPE[b] | Subjective and objective burden | Developed as part of WHO[c] mhGAP[d] and PRIME[e] | TPO[f] Nepal, a Nepali nongovernmental organization | 5 sessions | 4 days | Not specified | Not specified |
| Rajai et al [45] | 2021 | Peer- assisted education | Subjective and objective burden | Researchers and approved by 3 faculty members of the Army College of Medical Sciences | Trained peer facilitators | 6 sessions | 3 weeks | Twice weekly | 1 hour |
| Zhou et al [44] | 2020 | FLEP[g] | Subjective and objective burden | FLEP: peer-led psychoeducation program developed from the stress and coping model by Pearlin et al [70] | Group facilitators who were peer specialists of experienced caregivers | 8 sessions | 8 weeks | Once weekly | 2 hours |
| Dewi et al [37] | 2019 | Family psychoeducation and care decision without pasung | Subjective and objective burden | Researchers | Researchers | 3 sessions | 3 weeks | Not specified | 35-40 minutes |
| Tabeleão et al [36] | 2018 | Psychoeducation | Subjective and objective burden | Not specified | 10 psychologists | 6 sessions | Not specified | Not specified | Not specified |
| Ntsayagae [40] | 2017 | Psychoeducation | Subjective and objective burden | Researcher based on existing literature | Researcher | 2 sessions | Not specified | Not specified | Not specified |
| De Souza et al [39] | 2016 | Psychoeducation | Subjective and objective burden | Psychoeducational intervention for patients with BD[h] by Colom et al [71] translated and adapted by Dell-Aglio et al [72] | Psychiatrist responsible for specific case | 6 sessions | 6 to 8 weeks | Twice weekly | Not specified |
| Kolostoumpis et al [38] | 2015 | Psychoeducation | Subjective burden | Adapted from the treatment protocol developed by Reinares et al [73] in the Barcelona Bipolar Disorders Program in Spain | Psychiatrist and psychologist | 7 sessions | Not specified | Not specified | Not specified |
| Jessy and Kalaimathy [35] | 2014 | Psychoeducation | Subjective and objective burden | Researcher based on WHO guidelines for educational interventions and existing literature. Manual translated into Tamil. | Researcher | 5 modules | 10 weeks | Twice weekly | 1 hour |

| Study | Year of publication | Type of intervention | Type of burden targeted by the intervention | Intervention developer | Intervention deliverer | Number of sessions | Duration of intervention | Frequency of sessions per week | Length of the sessions |
|---|---|---|---|---|---|---|---|---|---|
| Fallahi Khoshknab et al [41] | 2014 | Group psychoeducation | Subjective and objective burden | Researchers based on educational program of psychiatric nursing and psychiatric textbooks | Organized based on educational program of psychiatric nursing and psychiatric textbooks (Campbell [74]) and converted to understandable text for patients and families | 4 sessions | 4 weeks | Once weekly | 2 hours |
| Navidian et al [42] | 2012 | Psychoeducation | Subjective burden | Researchers based on the families' needs and the existing literature | Psychiatrist and mental health nurse | 4 sessions | 4 weeks | Once weekly | 2 hours |
| Sharif et al [75] | 2012 | Psychoeducation | Subjective and objective burden | Psychiatrist and psychiatric nurse based on the literature and needs of the families | Psychiatrist, nurse, and guest speakers | 10 sessions | 5 weeks | Not specified | 1.5 hours |
| Chiu et al [43] | 2011 | FLEP | Subjective burden | 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[i] Family-to-Family program | Trainers who were themselves family members of people with SMI[j]. They received training and a trainer manual. | 8 sessions | Not specified | Once weekly | Not specified |
| Chien et al [48] | 2018 | Family-led support program | Subjective and objective burden | Family-led mutual support group—contents were based on similar program protocols and the researcher-developed family mutual support groups for psychotic disorders | Family-led mutual support group—co-led by 2 peer family caregivers along with a researcher and rehabilitation nurse | 16 sessions | 36 weeks | Twice weekly | 2 hours |
| Mentis et al [49] | 2015 | Support group intervention | Subjective burden | Nongovernmental mental health organization (NGOMH[k]) | Not mentioned | Not specified | Not specified | Not specified | Not specified |
| Chien et al [48] | 2018 | Family-led mutual support program | Subjective burden | 6 experts on psychiatric rehabilitation (including psychiatrists, clinical psychologists, and nurse specialists) | Peer leader who received training from the researchers and worked closely with a group leader who was a trained psychiatric nurse | 14 sessions | 39 weeks | Twice weekly | 2 hours |

| Study | Year of publication | Type of intervention | Type of burden targeted by the intervention | Intervention developer | Intervention deliverer | Number of sessions | Duration of intervention | Frequency of sessions per week | Length of the sessions |
|---|---|---|---|---|---|---|---|---|---|
| Chien and Chan [47] | 2013 | FPGP[l] | Subjective and objective burden | 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. | Peer support—peer leader supported by principal researcher | 14 group sessions | 39 weeks | Twice weekly | 2 hours |
| Behrouian et al [62] | 2021 | Emotional regulation | Subjective burden | On the basis of the Dialectical Behavior Therapy Skills Workbook and CBT[m] principles | Clinical psychologist | 8 sessions | 8 weeks | Once weekly | Not specified |
| Behrouian et al [63] | 2020 | Emotion regulation training | Subjective and objective burden | Sessions were based on previous studies (Gratz and Gunderson [77]). Trainings were based on the Dialectical Behavior Therapy Skills Workbook (McKay et al [78]). | Clinical psychologist | 8 sessions | Not specified | Not specified | Not specified |
| Faghih and Pahlavanzadeh [79] | 2019 | CBT | Subjective burden | Researchers based on existing literature | Researcher | 16 sessions | 8 weeks | Twice weekly | 1.5 hours |
| Khosravi et al [65] | 2021 | Spirituality-based program | Subjective burden | Researcher | Researcher | 6 sessions | 8 weeks | Twice weekly | 1 hour |
| Chien et al [58] | 2020 | Manual-guided PBSP[n] | Subjective and objective burden | Developed by McCann et al [80] in Australia, and its Chinese translated version was validated and refined by the research team. | Psychiatric nurse | 5 modules | 21 weeks | Not specified | Not specified |
| McCann et al [59] | 2015 | Guided self-help manual | Subjective and objective burden | The Good Mood Guide developed by Lifeline South Coast NSW[o], Australia | Researcher via telephone | 8 modules | 8 weeks | Once weekly | Not specified |
| Deb et al [67] | 2018 | Smartphone-based interventions (MHapps[p]) | Subjective and objective burden | Researcher | Psychiatrist | Not specified | Not specified | Not specified | Not specified |

| Study | Year of publication | Type of intervention | Type of burden targeted by the intervention | Intervention developer | Intervention deliverer | Number of sessions | Duration of intervention | Frequency of sessions per week | Length of the sessions |
|---|---|---|---|---|---|---|---|---|---|
| Hyun et al [61] | 2018 | IEP[q] | Subjective burden | Developed by Hyun et al [61] for community-living PMIs[r] based on the empowerment theories of Kanter [81] and McLean [82] | Mental health professionals who received training with a written structured intervention protocol from the research team | 4 sessions | 4 weeks | Once weekly | 2 hours |
| Hou et al [60] | 2014 | MBSR[s] program | Subjective burden | Not stated but assumed that the researchers modeled the program on the original MBSR by Kabat-Zinn [83] | 3 trained instructors with >3 years of experience in MBSR | 8 sessions | 8 weeks | Once weekly | 2 hours, with a CD of 30-45 minutes for home practice |
| Zoladl et al [66] | 2020 | Participatory care model | Subjective burden | Researchers | Researchers and staff at the research site | 8 sessions | 12 weeks | Once weekly | 1.5 hours |
| Ramirez et al [68] | 2017 | Standard care—included care from general medicine, psychiatry, psychology, family therapy, neuropsychological rehabilitation, and occupational therapy | Subjective and objective burden | Not stated | Medical officer, psychiatrist, psychologist, and occupational therapist | 12-18 sessions | 10 weeks | Not specified | Not specified |

[a]CoMHIP: community mental health early intervention project.

[b]RESHAPE: Reducing Stigma Among Health Care Providers.

[c]WHO: World Health Organization.

[d]mhGAP: Mental Health Gap Action Programme.

[e]PRIME: Programme for Improving Mental Health Care.

[f]TPO: Transcultural Psychosocial Organization.

[g]FLEP: Family Link Education Programme.

[h]BD: bipolar disorder.

[i]NAMI: National Alliance on Mental Illness.

[j]SMI: severe mental illness.

[k]NGOMH: nongovernmental mental health.

[l]FPGP: Family-led Peer Support Group Program.

[m]CBT: cognitive behavioral therapy.

[n]PBSP: problem-solving–based self-learning program.

[o]NSW: New South Wales.

[p]MHapps: mental health apps.

[q]IEP: integrated empowerment program.

[r]PMI: person with mental illness.

[s]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 Witwatersrand 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

1. Pillay AL. Is deinstitutionalisation a cheap alternative to chronic mental health care? S Afr J Psychol 2017 May 10;47(2):141-147 [FREE Full text] [doi: 10.1177/0081246317709959]
2. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
3. Roth DL, Fredman L, Haley WE. Informal caregiving and its impact on health: a reappraisal from population-based studies. Gerontologist 2015 Apr;55(2):309-319 [FREE Full text] [doi: 10.1093/geront/gnu177] [Medline: 26035608]
4. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178 [FREE Full text] [doi: 10.1016/s2215-0366(15)00505-2]
5. Lynch SH, Shuster G, Lobo ML. The family caregiver experience - examining the positive and negative aspects of compassion satisfaction and compassion fatigue as caregiving outcomes. Aging Ment Health 2018 Nov;22(11):1424-1431 [FREE Full text] [doi: 10.1080/13607863.2017.1364344] [Medline: 28812375]
6. Dixon LB, Glynn SM, Cohen AN, Drapalski AL, Medoff D, Fang LJ, et al. Outcomes of a brief program, REORDER, to promote consumer recovery and family involvement in care. Psychiatr Serv 2014 Jan 01;65(1):116-120 [FREE Full text] [doi: 10.1176/appi.ps.201300074] [Medline: 24177229]
7. Chadda RK. Caring for the family caregivers of persons with mental illness. Indian J Psychiatry 2014 Jul;56(3):221-227 [FREE Full text] [doi: 10.4103/0019-5545.140616] [Medline: 25316932]
8. Paruk S, Ramdhial M. Prevalence of caregiver burden, depressive and anxiety symptoms in caregivers of children with psychiatric disorders in Durban, South Africa. S Afr J Psychiatr 2018 Sep 20;24:a1314 [FREE Full text] [doi: 10.4102/sajpsychiatry.v24i0.1314]
9. Naughton-Doe R, Moran N, Wakeman E, Wilberforce M, Bennett L, Webber M. Interventions that support unpaid carers of adult mental health inpatients: a scoping review. J Ment Health 2022 May 09:1-17 [FREE Full text] [doi: 10.1080/09638237.2022.2069702] [Medline: 35532039]
10. Yesufu-Udechuku A, Harrison B, Mayo-Wilson E, Young N, Woodhams P, Shiers D, et al. Interventions to improve the experience of caring for people with severe mental illness: systematic review and meta-analysis. Br J Psychiatry 2015 Apr;206(4):268-274 [FREE Full text] [doi: 10.1192/bjp.bp.114.147561] [Medline: 25833867]
11. Chisholm D, Docrat S, Abdulmalik J, Alem A, Gureje O, Gurung D, et al. Mental health financing challenges, opportunities and strategies in low- and middle-income countries: findings from the Emerald project. BJPsych Open 2019 Aug 06;5(5):e68 [FREE Full text] [doi: 10.1192/bjo.2019.24] [Medline: 31530327]
12. Alexander G, Bebee CE, Chen KM, Vignes RM, Dixon B, Escoffery R, et al. Burden of caregivers of adult patients with schizophrenia in a predominantly African ancestry population. Qual Life Res 2016 Feb;25(2):393-400 [FREE Full text] [doi: 10.1007/s11136-015-1077-5] [Medline: 26205769]

13. Mokgothu MC, Du Plessis E, Koen MP. The strengths of families in supporting mentally-ill family members. Curationis 2015 Apr 10;38(1):1258 [FREE Full text] [doi: 10.4102/curationis.v38i1.1258] [Medline: 26018195]
14. Kimotho SG. Understanding the nature of stigma communication associated with mental illness in Africa: a focus on cultural beliefs and stereotypes. In: Deconstructing Stigma in Mental Health. Hershey, PA: IGI Global; 2018.
15. Egbe CO, Brooke-Sumner C, Kathree T, Selohilwe O, Thornicroft G, Petersen I. Psychiatric stigma and discrimination in South Africa: perspectives from key stakeholders. BMC Psychiatry 2014 Jul 4;14(1):191 [FREE Full text] [doi: 10.1186/1471-244x-14-191]
16. Yerriah J, Tomita A, Paruk S. Surviving but not thriving: burden of care and quality of life for caregivers of patients with schizophrenia spectrum disorders and comorbid substance use in South Africa. Early Interv Psychiatry 2022 Feb;16(2):153-161. [doi: 10.1111/eip.13141] [Medline: 33733599]
17. Kayaalp A, Page KJ, Rospenda KM. Caregiver burden, work-family conflict, family-work conflict, and mental health of caregivers: a mediational longitudinal study. Work Stress 2021;35(3):217-240 [FREE Full text] [doi: 10.1080/02678373.2020.1832609] [Medline: 34483432]
18. Ewertzon M, Hanson E. Support interventions for family members of adults with mental illness: a narrative literature review. Issues Ment Health Nurs 2019 Sep;40(9):768-780. [doi: 10.1080/01612840.2019.1591547] [Medline: 31145025]
19. Soo SA, Zhang ZW, Khong SJ, Low JE, Thambyrajah VS, Alhabsyi SH, et al. Randomized controlled trials of psychoeducation modalities in the management of bipolar disorder: a systematic review. J Clin Psychiatry 2018;79(3):17r11750. [doi: 10.4088/JCP.17r11750] [Medline: 29727072]
20. Day K, Petrakis M. Family interventions in early psychosis service delivery: a systematized review. Soc Work Ment Health 2017;15(6):632-650 [FREE Full text] [doi: 10.1080/15332985.2016.1271381]
21. Claxton M, Onwumere J, Fornells-Ambrojo M. Do family interventions improve outcomes in early psychosis? A systematic review and meta-analysis. Front Psychol 2017 Mar 27;8:371 [FREE Full text] [doi: 10.3389/fpsyg.2017.00371] [Medline: 28396643]
22. Bademli K, Cetinkaya Duman Z. Family to family support programs for the caregivers of schizophrenia patients: a systematic review. Turk Psikiyatri Derg 2011;22(4):255-265 [FREE Full text] [Medline: 22143951]
23. Chen LF, Liu J, Zhang J, Lu XQ. Non-pharmacological interventions for caregivers of patients with schizophrenia: a meta-analysis. Psychiatry Res 2016 Jan 30;235:123-127 [FREE Full text] [doi: 10.1016/j.psychres.2015.11.037] [Medline: 26639649]
24. Sin J, Gillard S, Spain D, Cornelius V, Chen T, Henderson C. Effectiveness of psychoeducational interventions for family carers of people with psychosis: a systematic review and meta-analysis. Clin Psychol Rev 2017 Aug;56:13-24 [FREE Full text] [doi: 10.1016/j.cpr.2017.05.002] [Medline: 28578249]
25. Napa W, Tungpunkom P, Pothimas N. Effectiveness of family interventions on psychological distress and expressed emotion in family members of individuals diagnosed with first-episode psychosis: a systematic review. JBI Database Syst Rev Implement Rep 2017 Apr;15(4):1057-1079. [doi: 10.11124/JBISRIR-2017-003361] [Medline: 28398985]
26. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
27. Sin J, Henderson C, Woodham LA, Sesé Hernández A, Gillard S. A multicomponent eHealth intervention for family carers for people affected by psychosis: a coproduced design and build study. J Med Internet Res 2019 Aug 06;21(8):e14374 [FREE Full text] [doi: 10.2196/14374] [Medline: 31389333]
28. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-based interventions to improve mental health, general caregiving outcomes, and general health for informal caregivers of adults with chronic conditions living in the community: rapid evidence review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]
29. Silaule O, Nkosi NG, Adams F. Extent of caregiver burden among informal caregivers of persons with severe mental disorders in rural South Africa. Rural Remote Health 2023 Jun;23(2):7509 [FREE Full text] [doi: 10.22605/RRH7509] [Medline: 37264594]
30. Peters MD, Godfrey C, McInerney P, Munn Z, Trico AC, Khalil H. Chapter 11: scoping reviews. In: Aromataris E, Munn Z, editors. Joanna Briggs Institute Reviewer's Manual. Adelaide, Australia: The Joanna Briggs Institute; 2020.
31. Silaule O, Adams F, Nkosi NG. Strategies for alleviating the burden experienced by informal caregivers of persons with severe mental disorders in transitional countries: protocol for a scoping review. JMIR Res Protoc 2023 Jul 24;12:e44268 [FREE Full text] [doi: 10.2196/44268] [Medline: 37486756]
32. Liu Z, Heffernan C, Tan J. Caregiver burden: a concept analysis. Int J Nurs Sci 2020 Jul 25;7(4):438-445 [FREE Full text] [doi: 10.1016/j.ijnss.2020.07.012] [Medline: 33195757]
33. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
34. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 09;16:15 [FREE Full text] [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]

35. Jessy MF. Empowerment of caregivers of persons with schizophrenia: a group psychoeducational intervention. Mother Teresa Womens University. 2014. URL: https://shodhganga.inflibnet.ac.in:8443/jspui/handle/10603/289368 [accessed 2023-11-27]
36. Tabeleão V, Tomasi E, de Avila Quevedo L. A randomized, controlled trial of the effectiveness of a psychoeducational intervention on family caregivers of patients with mental disorders. Community Ment Health J 2018 Feb 25;54(2):211-217. [doi: 10.1007/s10597-017-0126-7] [Medline: 28444574]
37. Dewi AR, Daulima NH, Wardani IY. Managing family burden through combined family psychoeducation and care decision without pasung therapies. Enfermería Clínica 2019 Sep;29:76-80. [doi: 10.1016/j.enfcli.2019.04.012]
38. Kolostoumpis D, Bergiannaki JD, Peppou LE, Louki E, Fousketaki S, Patelakis A, et al. Effectiveness of relatives' psychoeducation on family outcomes in bipolar disorder. Int J Ment Health 2015 Sep 15;44(4):290-302. [doi: 10.1080/00207411.2015.1076292]
39. de Souza MS, da Silva RA, Molina MA, Jansen K, de Lima Ferreira L, Kelbert EF, et al. Six-session caregiver psychoeducation on bipolar disorder: does it bring benefits to caregivers? Int J Soc Psychiatry 2016 Jun 14;62(4):377-385. [doi: 10.1177/0020764016636909] [Medline: 26975694]
40. Ntsayagae EI. A psycho-educational programme for caregivers of mentally ill persons to facilitate their mental health. University of Johannesburg. 2017 Oct. URL: https://www.proquest.com/openview/cfa667c9ca773bc0b59f3df3d6d863f5/1?pq-origsite=gscholar&cbl=2026366&diss=y [accessed 2023-11-28]
41. Fallahi Khoshknab M, Sheikhona M, Rahgouy A, Rahgozar M, Sodagari F. The effects of group psychoeducational programme on family burden in caregivers of Iranian patients with schizophrenia. J Psychiatr Ment Health Nurs 2014 Jun 26;21(5):438-446. [doi: 10.1111/jpm.12107] [Medline: 23980535]
42. Navidian A, Kermansaravi F, Rigi SN. The effectiveness of a group psycho-educational program on family caregiver burden of patients with mental disorders. BMC Res Notes 2012 Aug 01;5:399 [FREE Full text] [doi: 10.1186/1756-0500-5-399] [Medline: 22853873]
43. Chiu MY, Wei GF, Lee S, Choovanichvong S, Wong FH. Empowering caregivers: impact analysis of FamilyLink Education Programme (FLEP) in Hong Kong, Taipei and Bangkok. Int J Soc Psychiatry 2013 Feb;59(1):28-39 [FREE Full text] [doi: 10.1177/0020764011423171] [Medline: 21971981]
44. Zhou DH, Chiu YL, Lo TL, Lo WF, Wong SS, Luk KL, et al. Outside-in or inside-out? A randomized controlled trial of two empowerment approaches for family caregivers of people with schizophrenia. Issues Ment Health Nurs 2020 Sep 04;41(9):761-772. [doi: 10.1080/01612840.2020.1734992] [Medline: 32497453]
45. Rajai N, Lami B, Pishgooie AH, Habibi H, Alavizerang F. Evaluating the effect of peer-assisted education on the functioning in family caregivers of patients with schizophrenia: a clinical trial study. Korean J Fam Med 2021 Sep;42(5):356-362 [FREE Full text] [doi: 10.4082/kjfm.20.0098] [Medline: 34607410]
46. Stanley S, Balakrishnan S, Ilangovan S. Correlates of caregiving burden in schizophrenia: a cross-sectional, comparative analysis from India. Soc Work Ment Health 2017;15(3):284-307 [FREE Full text] [doi: 10.1080/15332985.2016.1220440]
47. Chien WT, Chan SW. The effectiveness of mutual support group intervention for Chinese families of people with schizophrenia: a randomised controlled trial with 24-month follow-up. Int J Nurs Stud 2013 Oct;50(10):1326-1340. [doi: 10.1016/j.ijnurstu.2013.01.004] [Medline: 23433723]
48. Chien WT, Bressington D, Chan SW. A randomized controlled trial on mutual support group intervention for families of people with recent-onset psychosis: a four-year follow-up. Front Psychiatry 2018 Dec 18;9:710 [FREE Full text] [doi: 10.3389/fpsyt.2018.00710] [Medline: 30618880]
49. Mentis M, Messinis L, Kotrotsiou E, Angelopoulos NV, Marneras C, Papathanasopoulos P, et al. Efficacy of a support group intervention on psychopathological characteristics among caregivers of psychotic patients. Int J Soc Psychiatry 2015 Jun;61(4):373-378. [doi: 10.1177/0020764014547075] [Medline: 25149101]
50. Seshadri K, Sivakumar T, Jagannathan A. The family support movement and schizophrenia in India. Curr Psychiatry Rep 2019 Sep 14;21(10):95. [doi: 10.1007/s11920-019-1081-5] [Medline: 31522258]
51. Sagut P, Çetinkaya Duman Z. Comparison of caregiver burden in first episode versus chronic psychosis. Arch Psychiatr Nurs 2016 Dec;30(6):768-773 [FREE Full text] [doi: 10.1016/j.apnu.2016.07.011] [Medline: 27888973]
52. AlMakhamreh SS. Exploring experiences of informal carers of mental health: developing community intervention in social work in Jordan. Int Soc Work 2017 Mar 16;61(6):1042-1053 [FREE Full text] [doi: 10.1177/0020872817695391]
53. Bonsu AS, Salifu Yendork J, Teye-Kwadjo E. The influence of caregiver stress and affiliate stigma in community-based mental health care on family caregiver wellbeing. Int J Ment Health 2020;49(1):35-55 [FREE Full text] [doi: 10.1080/00207411.2020.1717053]
54. Wang M, Jin G, Wei Y, Wang F, Pan Z, Chen L, et al. Awareness, utilization and influencing factors of social supports for main informal caregivers of schizophrenia patients: a cross-sectional study in primary care settings in Beijing, China. BMC Fam Pract 2020 Sep 17;21(1):192 [FREE Full text] [doi: 10.1186/s12875-020-01257-z] [Medline: 32943006]
55. Rai S, Gurung D, Kaiser BN, Sikkema KJ, Dhakal M, Bhardwaj A, et al. A service user co-facilitated intervention to reduce mental illness stigma among primary healthcare workers: utilizing perspectives of family members and caregivers. Fam Syst Health 2018 Jun;36(2):198-209 [FREE Full text] [doi: 10.1037/fsh0000338] [Medline: 29902036]

56. Ng P, Young DK, Pan J, Law KK. The effects of a community mental health intervention project upon family caregivers of persons with suspected mental illness in the Chinese cultural context. Int Soc Work 2017 Mar 20;61(6):1067-1082. [doi: 10.1177/0020872817695642]
57. Chen L, Zhao Y, Tang J, Jin G, Liu Y, Zhao X, et al. The burden, support and needs of primary family caregivers of people experiencing schizophrenia in Beijing communities: a qualitative study. BMC Psychiatry 2019 Feb 20;19(1):75 [FREE Full text] [doi: 10.1186/s12888-019-2052-4] [Medline: 30786852]
58. Chien WT, Bressington D, Lubman DI, Karatzias T. A randomised controlled trial of a caregiver-facilitated problem-solving based self-learning program for family carers of people with early psychosis. Int J Environ Res Public Health 2020 Dec 14;17(24):9343 [FREE Full text] [doi: 10.3390/ijerph17249343] [Medline: 33327452]
59. McCann TV, Songprakun W, Stephenson J. A randomized controlled trial of guided self-help for improving the experience of caring for carers of clients with depression. J Adv Nurs 2015 Jul;71(7):1600-1610. [doi: 10.1111/jan.12624] [Medline: 25656334]
60. Hou RJ, Wong SY, Yip BH, Hung AT, Lo HH, Chan PH, et al. The effects of mindfulness-based stress reduction program on the mental health of family caregivers: a randomized controlled trial. Psychother Psychosom 2014;83(1):45-53. [doi: 10.1159/000353278] [Medline: 24281411]
61. Hyun MS, Nam KA, Kim H. Effects of a brief empowerment program for families of persons with mental illness in South Korea: a pilot study. Issues Ment Health Nurs 2019 Jan;40(1):8-14 [FREE Full text] [doi: 10.1080/01612840.2018.1453000] [Medline: 29847184]
62. Behrouian M, Ramezani T, Dehghan M, Sabahi A, Ebrahimnejad Zarandi B. The effect of the emotion regulation training on the resilience of caregivers of patients with schizophrenia: a parallel randomized controlled trial. BMC Psychol 2021 Mar 02;9(1):39 [FREE Full text] [doi: 10.1186/s40359-021-00542-5] [Medline: 33653410]
63. Behrouian M, Ramezani T, Dehghan M, Sabahi A, Ebrahimnejad Zarandi B. The effect of emotion regulation training on stress, anxiety, and depression in family caregivers of patients with schizophrenia: a randomized controlled trial. Community Ment Health J 2020 Aug;56(6):1095-1102 [FREE Full text] [doi: 10.1007/s10597-020-00574-y] [Medline: 32056061]
64. Souza AL, Guimarães RA, de Araújo Vilela D, de Assis RM, de Almeida Cavalcante Oliveira LM, Souza MR, et al. Factors associated with the burden of family caregivers of patients with mental disorders: a cross-sectional study. BMC Psychiatry 2017 Oct 25;17(1):353 [FREE Full text] [doi: 10.1186/s12888-017-1501-1] [Medline: 29070012]
65. Khosravi F, Fereidooni-Moghadam M, Mehrabi T, Moosavizade SR. The effect of a spirituality-based program on stress, anxiety, and depression of caregivers of patients with mental disorders in Iran. J Relig Health 2022 Feb;61(1):93-108 [FREE Full text] [doi: 10.1007/s10943-021-01372-w] [Medline: 34417681]
66. Zoladl M, Afroughi S, Nooryan K, Kharamin S, Haghgoo A, Parandvar Y. Applying collaborative care model on intensive caregiver burden and resilient family caregivers of patients with mental disorders: a randomized controlled trial. Iran J Psychiatry 2020 Feb 25. [doi: 10.18502/ijps.v15i1.2436]
67. Sinha Deb K, Tuli A, Sood M, Chadda R, Verma R, Kumar S, et al. Is India ready for mental health apps (MHApps)? A quantitative-qualitative exploration of caregivers' perspective on smartphone-based solutions for managing severe mental illnesses in low resource settings. PLoS One 2018 Sep 19;13(9):e0203353 [FREE Full text] [doi: 10.1371/journal.pone.0203353] [Medline: 30231056]
68. Ramírez A, Palacio JD, Vargas C, Díaz-Zuluaga AM, Duica K, Agudelo Berruecos Y, et al. Expressed emotions, burden and family functioning in schizophrenic and bipolar I patients of a multimodal intervention programme: PRISMA. Revista Colombiana de Psiquiatría (English ed.) 2017 Jan;46(1):2-11 [FREE Full text] [doi: 10.1016/j.rcpeng.2017.03.008]
69. Zeng Y, Zhou Y, Lin J. Perceived burden and quality of life in Chinese caregivers of people with serious mental illness: a comparison cross-sectional survey. Perspect Psychiatr Care 2017 Jul 23;53(3):183-189. [doi: 10.1111/ppc.12151] [Medline: 26909775]
70. Pearlin LI, Mullan JT, Semple SJ, Skaff MM. Caregiving and the stress process: an overview of concepts and their measures. Gerontologist 1990 Oct;30(5):583-594. [doi: 10.1093/geront/30.5.583] [Medline: 2276631]
71. Colom F, Vieta E, Scott J. Psychoeducation Manual for Bipolar Disorder. Cambridge, MA: Cambridge University Press; 2010.
72. Dell'Aglio JCJ, Figueoredo AL, Souza LD, Argimon IL. Cognitive behavioral model of bipolar disorder. In: Andretta I, Oliveira MS, editors. Pratical Manual of Cognitive Behavioral Therapy. São Paulo, Brazil: Casa do Psicólogo; 2011.
73. Reinares M, Colom F, Rosa AR, Bonnín CM, Franco C, Solé B, et al. The impact of staging bipolar disorder on treatment outcome of family psychoeducation. J Affect Disord 2010 Jun;123(1-3):81-86. [doi: 10.1016/j.jad.2009.09.009] [Medline: 19853922]
74. Campbell RJ. Campbell's Psychiatric Dictionary, 9th edition. New York, NY: Oxford University Press; 2009.
75. Sharif F, Shaygan M, Mani A. Effect of a psycho-educational intervention for family members on caregiver burdens and psychiatric symptoms in patients with schizophrenia in Shiraz, Iran. BMC Psychiatry 2012 May 25;12(1):48 [FREE Full text] [doi: 10.1186/1471-244x-12-48]
76. Li Z, Arthur D. Family education for people with schizophrenia in Beijing, China: randomised controlled trial. Br J Psychiatry 2005 Oct;187:339-345. [doi: 10.1192/bjp.187.4.339] [Medline: 16199793]

77. Gratz KL, Gunderson JG. Preliminary data on an acceptance-based emotion regulation group intervention for deliberate self-harm among women with borderline personality disorder. Behav Ther 2006 Mar;37(1):25-35. [doi: 10.1016/j.beth.2005.03.002] [Medline: 16942958]
78. McKay M, Wood JC. The Dialectical Behavior Therapy Skills Workbook: Practical Dbt Exercises for Learning Mindfulness, Interpersonal Effectiveness, Emotion Regulation, Emotion Regulation, and Distress Tolerance. Oakland, CA: New Harbinger Publications; 2007.
79. Faghih M, Pahlavanzadeh S. The effect of cognitive behavioral therapy on the burden in drug dependent persons' caregivers: a randomized controlled clinical trial. Iran J Nurs Midwifery Res 2019;24(2):131-136 [FREE Full text] [doi: 10.4103/ijnmr.IJNMR_86_18] [Medline: 30820225]
80. McCann TV, Cotton SM, Lubman DI. Social problem solving in carers of young people with a first episode of psychosis: a randomized controlled trial. Early Interv Psychiatry 2017 Aug;11(4):346-350 [FREE Full text] [doi: 10.1111/eip.12301] [Medline: 26592195]
81. Kanter RM. Men and Women of the Corporation. New York, NY: Basic Books; 1993.
82. McLean A. Empowerment and the psychiatric consumer/ex-patient movement in the United States: contradictions, crisis and change. Soc Sci Med 1995 Apr;40(8):1053-1071. [doi: 10.1016/0277-9536(94)00179-w] [Medline: 7597459]
83. Kabat-Zinn J, Hanh TN. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York, NY: Delacorte Press; 1990.
84. Olivier B, Verdonck M, Caseleijn D. Digital technologies in undergraduate and postgraduate education in occupational therapy and physiotherapy: a scoping review. JBI Evid Synth 2020 May;18(5):863-892. [doi: 10.11124/JBISRIR-D-19-00210] [Medline: 32813350]
85. Ayalew M, Workicho A, Tesfaye E, Hailesilasie H, Abera M. Burden among caregivers of people with mental illness at Jimma University Medical Center, Southwest Ethiopia: a cross-sectional study. Ann Gen Psychiatry 2019 Jun 24;18:10 [FREE Full text] [doi: 10.1186/s12991-019-0233-7] [Medline: 31285750]
86. Udoh EE, Omorere DE, Sunday O, Osasu OS, Amoo BA. Psychological distress and burden of care among family caregivers of patients with mental illness in a neuropsychiatric outpatient clinic in Nigeria. PLoS One 2021 May 6;16(5):e0250309 [FREE Full text] [doi: 10.1371/journal.pone.0250309] [Medline: 33956799]
87. Marimbe BD, Cowan F, Kajawu L, Muchirahondo F, Lund C. Perceived burden of care and reported coping strategies and needs for family caregivers of people with mental disorders in Zimbabwe. Afr J Disabil 2016 Aug 24;5(1):209 [FREE Full text] [doi: 10.4102/ajod.v5i1.209] [Medline: 28730046]
88. National mental health strategy 2012/13 - 2015/16. Federal Democratic Republic of Ethiopia Ministry of Health. URL: http://repository.iifphc.org/bitstream/handle/123456789/264/104.12%20National%20Mental%20Health%20Strategy%20--%20ETHIOPIA.pdf?sequence=1&isAllowed=y [accessed 2023-11-28]
89. Caregivers and mental illness: living with Schizophrenia. World Federation for Mental Health (WFMH). 2014. URL: http://discovery.ucl.ac.uk/137713/ [accessed 2023-11-28]
90. The world health report, 2001: mental health; new understanding, new hope. World Health Organization. 2001. URL: https://unesdoc.unesco.org/ark:/48223/pf0000123325 [accessed 2023-11-28]
91. Reupert A, Ward B, McCormick F, Ward C, Waller S, Kidd S. Developing a model of family focused practice with consumers, families, practitioners and managers: a community based participatory research approach. BMC Health Serv Res 2018 Jan 30;18(1):66 [FREE Full text] [doi: 10.1186/s12913-018-2844-0] [Medline: 29382350]
92. Maybery D, Jaffe IC, Cuff R, Duncan Z, Grant A, Kennelly M, et al. Mental health service engagement with family and carers: what practices are fundamental? BMC Health Serv Res 2021 Oct 09;21(1):1073 [FREE Full text] [doi: 10.1186/s12913-021-07104-w] [Medline: 34627245]
93. Cameron SL, Tchernegovski P, Maybery D. Mental health service users' experiences and perspectives of family involvement in their care: a systematic literature review. J Ment Health 2023 Jun;32(3):699-715 [FREE Full text] [doi: 10.1080/09638237.2022.2091760] [Medline: 35808821]
94. Olawale KO, Mosaku KS, Fatoye FO, Mapayi BM, Oginni OA. Caregiver burden in families of patients with depression attending Obafemi Awolowo University teaching hospitals complex Ile-Ife Nigeria. Gen Hosp Psychiatry 2014;36(6):743-747 [FREE Full text] [doi: 10.1016/j.genhosppsych.2014.08.008] [Medline: 25217492]
95. Rathod S, Pinninti N, Irfan M, Gorczynski P, Rathod P, Gega L, et al. Mental health service provision in low- and middle-income countries. Health Serv Insights 2017 Mar 28;10:1178632917694350 [FREE Full text] [doi: 10.1177/1178632917694350] [Medline: 28469456]
96. Ojha SP, Pant SB. Community based rehabilitation for task shifting in mental health. J Psychosoc Rehabil Ment Health 2018 Nov 29;5(2):109-110 [FREE Full text] [doi: 10.1007/s40737-018-0128-y]
97. Deimling Johns L, Power J, MacLachlan M. Community-based mental health intervention skills: task shifting in low- and middle-income settings. Int Perspect Psychol 2018 Oct;7(4):205-230. [doi: 10.1037/ipp0000097]
98. Madigan K, Egan P, Brennan D, Hill S, Maguire B, Horgan F, et al. A randomised controlled trial of carer-focussed multi-family group psychoeducation in bipolar disorder. Eur Psychiatry 2012 May;27(4):281-284 [FREE Full text] [doi: 10.1016/j.eurpsy.2010.12.008] [Medline: 21334858]

99. Ebrahimi H, Seyedfatemi N, Namdar Areshtanab H, Ranjbar F, Thornicroft G, Whitehead B, et al. Barriers to family caregivers' coping with patients with severe mental illness in Iran. Qual Health Res 2018 May 24;28(6):987-1001. [doi: 10.1177/1049732318758644] [Medline: 29478404]
100. Braathen SH, Vergunst R, Mji G, Mannan H, Swartz L. Understanding the local context for the application of global mental health: a rural South African experience. Int Health 2013 Mar;5(1):38-42. [doi: 10.1093/inthealth/ihs016] [Medline: 24029844]
101. Javadi D, Feldhaus I, Mancuso A, Ghaffar A. Applying systems thinking to task shifting for mental health using lay providers: a review of the evidence. Glob Ment Health (Camb) 2017 Jul 31;4:e14 [FREE Full text] [doi: 10.1017/gmh.2017.15] [Medline: 29230310]
102. Ae-Ngibise KA, Doku VC, Asante KP, Owusu-Agyei S. The experience of caregivers of people living with serious mental disorders: a study from rural Ghana. Glob Health Action 2015 May 11;8(1):26957 [FREE Full text] [doi: 10.3402/gha.v8.26957] [Medline: 25967587]
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]
104. Adeleye OA, Ofili AN. Strengthening intersectoral collaboration for primary health care in developing countries: can the health sector play broader roles? J Environ Public Health 2010;2010:272896 [FREE Full text] [doi: 10.1155/2010/272896] [Medline: 20454703]
105. Woo BF, Tam WW, Williams MY, Ow Yong JQ, Cheong ZY, Ong YC, et al. Characteristics, methodological, and reporting quality of scoping reviews published in nursing journals: a systematic review. J Nurs Scholarsh 2023 Jul 09;55(4):874-885. [doi: 10.1111/jnu.12861] [Medline: 36494752]


## 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


*Edited by T de Azevedo Cardoso; submitted 29.04.23; peer-reviewed by DM Ramafikeng, Y Goh; comments to author 12.07.23; revised version received 20.08.23; accepted 04.10.23; published 18.01.24.*

*Please cite as:*
*Silaule O, Casteleijn D, Adams F, Nkosi NG*
*Strategies to Alleviate the Burden Experienced by Informal Caregivers of Persons With Severe Mental Disorders in Low- and Middle-Income Countries: Scoping Review*
*Interact J Med Res 2024;13:e48587*
*URL: https://www.i-jmr.org/2024/1/e48587*
*doi:10.2196/48587*
*PMID:38236636*

Review

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

Susanna Y Park[1*], PhD; Bridgette Do[2*], MPH, PhD; Jacqlyn Yourell[2], MS, PhD; Janice Hermer[3], MA; Jennifer Huberty[2], PhD

[1]Skylight, Radiant Foundation, Salt Lake City, UT, United States

[2]Fit Minded Inc, Phoenix, AZ, United States

[3]Arizona 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).

***(Interact J Med Res 2024;13:e48929)*** doi:10.2196/48929

**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,” “transcenden*,” and “life purpose” or “existential needs,” and a final search was performed on November 7, 2023, using the additional keywords of “spiritu*” and “relig*.” 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.

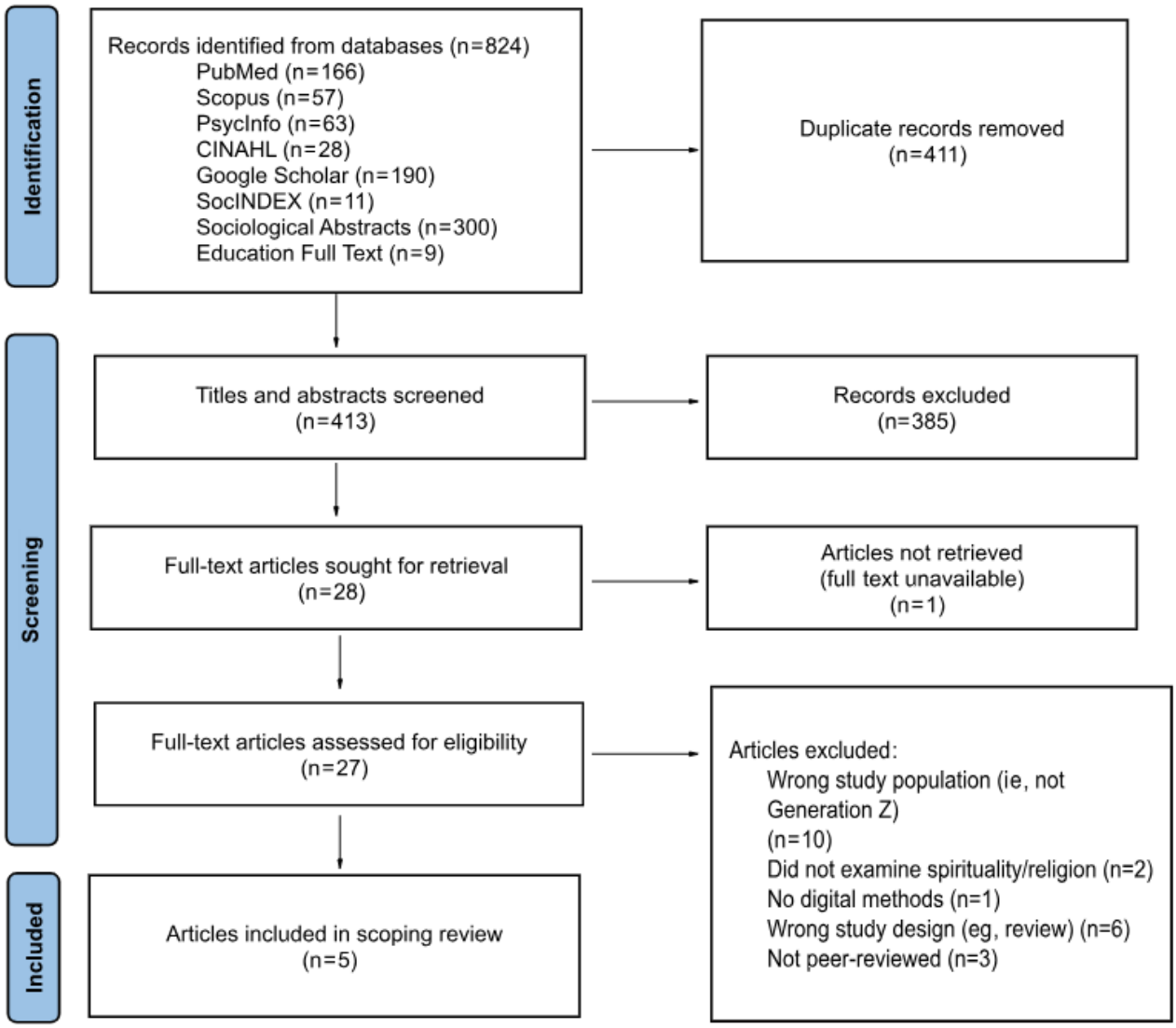


**Table 1.** Characteristics and main findings of the included studies.

| Reference and study country | Study design | Sample characteristics | Digital aspect[a] | Religion/spirituality measures[b] | Mental health measures[b] | Main findings |
|---|---|---|---|---|---|---|
| Mindu et al [25], 2023 (South Africa) | Mixed-methods study | 93 youth and young adults aged 16-24 years; 44% female | Assessed participants' knowledge and preferences for a digital mental health app | 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) | Mental health awareness (eg, taught about mental health and no prior education on mental health)<br>Mental health conditions affecting youth in the community (eg, substance abuse or misuse, posttraumatic stress disorder, depression, and anger) | • No participants had experience using a mental health app, but 99% indicated mental apps are important and can benefit youth.<br>• Religious and cultural beliefs were a barrier to using digital platforms (eg, social media). |
| Gao et al [26], 2021 (China) | Cross-sectional study | 1017 first-year college students (mean age 19 years); 77.8% female | Participants were recruited through an eHealth application to complete a survey | Religion (Response options: No religion; Buddhism; Christian; Others) | Generalized Anxiety Disorder-7 (GAD-7) and depression; Patient Health Questionnaire-9 (PHQ-9) [27,28] | • 95.3% indicated having no religion.<br>• Belief in Christianity and in Buddhism were associated with greater anxiety. |
| Craig Rushing et al [29], 2021 (United States) | Randomized controlled trial | 833 American Indian or Alaska Native teenagers and young adults aged 15-24 years; 66.3% female | 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) | Rate your spiritual health (Response options: Excellent; Very good; Good; Fair) [30] | Rate your mental health (Response options: Excellent; Very good; Good; Fair) [30] | • No significant differences between the 2 intervention arms within subjects.<br>• Mean scores of perceived health (physical, mental, and spiritual) significantly increased over time for both intervention arms.<br>• Those who reported better health also reported greater cultural resilience, identity, and cultural pride. |
| Reed et al [23], 2022 (United States) | Cross-sectional study | 349 American Indian or Alaska Native youth aged 15-24 years; 71.1% female | Assessed participants' use of media technologies (ie, media types, frequency, and duration) and how they use media technologies (ie, online behaviors and activities) | Select the top 3 health topics from a list of 15, including spiritual health | Self-reported mental health (How good is your mental health? Response options: 4-point Likert-type scale; 4=Excellent and 1=Poor)<br>Select the top 3 health topics from a list of 15, including mental health | • 53.5% of participants relied heavily on the internet to access health information.<br>• Nonsexual and gender minority participants reported better mental health than sexual and gender minority participants.<br>• The top 3 most important health topics were Native identity, mental health, and social justice and equality.<br>• Spiritual health was selected as the most important health topic 14% of the time. |

| Reference and study country | Study design | Sample characteristics | Digital aspect[a] | Religion/spirituality measures[b] | Mental health measures[b] | Main findings |
|---|---|---|---|---|---|---|
| Au-Yeung et al [24], 2023 (Canada) | Mixed-methods study | 5 Indigenous youth aged 11-16 years; 60.0% female | Assessed participants' opinions on the JoyPop app, and the accessibility and feasibility of the app | Qualitative question: "What does a 'Good Mind' (Haudenosaunee concept) mean to you?"<br>"The Good Mind is a physical, psychological, and spiritual journey that includes a reflective awareness of thoughts and intentions, and a way of being that is expressed through self-compassion and compassion for other beings." | Perceived mental wellness (5-point Likert scale)<br>Qualitative question: "Would you please describe what mental wellness means to you?" | • Participants had mixed ratings for their mental health from "fair" to "very good."<br>• Participants reported nature to be important for a "Good Mind."<br>• Participants reported enjoying the app's features (eg, games and breathing techniques) and made suggestions to make it relevant to their culture. |

[a]Refers to the digital component of the study, which may include the mode through which an intervention was given (eg, mobile text messages) or the constructs measured (eg, social media use and preferences for a mobile app).

[b]Measures refer to assessments, tools, or specific survey items that were used in the study to assess the construct of interest (eg, religion or spirituality, and mental health).

## 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, personalizable 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 1 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


1. Defining generations: Where Millennials end and Generation Z begins. Pew Research Center. URL: https://www.pewresearch.org/short-reads/2019/01/17/where-millennials-end-and-generation-z-begins/ [accessed 2024-01-25]
2. Katz R, Ogilvie S, Shaw J, Woodhead L. Gen Z, Explained: The Art of Living in a Digital Age. Chicago, IL: The University of Chicago Press; 2021.
3. State of Gen Z Mental Health. Harmony Healthcare IT. 2022. URL: https://www.harmonyhit.com/state-of-gen-z-mental-health/ [accessed 2024-01-25]
4. Gen Z Is Extremely Online. Morning Consult. 2022. URL: https://pro.morningconsult.com/instant-intel/gen-z-social-media-usage [accessed 2023-05-10]
5. Twenge JM. iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy–and Completely Unprepared for Adulthood–and What That Means for the Rest of Us. New York, NY: Atria Books; 2017.
6. Pichler S, Kohli C, Granitz N. DITTO for Gen Z: A framework for leveraging the uniqueness of the new generation. Business Horizons 2021 Sep;64(5):599-610. [doi: 10.1016/j.bushor.2021.02.021]
7. Duffy ME, Twenge JM, Joiner TE. Trends in mood and anxiety symptoms and suicide-related outcomes among U.S. undergraduates, 2007-2018: Evidence from two national surveys. J Adolesc Health 2019 Nov;65(5):590-598. [doi: 10.1016/j.jadohealth.2019.04.033] [Medline: 31279724]
8. Gen Z more likely to report mental health concerns. APA. URL: https://www.apa.org/monitor/2019/01/gen-z [accessed 2023-04-11]
9. Arrey AE, Bilsen J, Lacor P, Deschepper R. Spirituality/religiosity: A cultural and psychological resource among Sub-Saharan African migrant women with HIV/AIDS in Belgium. PLoS One 2016 Jul 22;11(7):e0159488 [FREE Full text] [doi: 10.1371/journal.pone.0159488] [Medline: 27447487]
10. Jaysawal N, Saha S. COVID-19 and spiritual well-being: implications for social work. Journal of Religion & Spirituality in Social Work: Social Thought 2022 Nov 18;42(2):135-151. [doi: 10.1080/15426432.2022.2127391]
11. Springtide Research Institute. The State of Religion & Young People 2022: Mental Health–What Faith Leaders Need to Know. Winona, MN: Springtide Research Institute; 2022.
12. Faith Like a Roundabout: Religion and Gen Z in an Uncertain World. Patheos. 2022. URL: https://www.patheos.com/blogs/religionnow/2022/01/faith-like-a-roundabout-religion-and-gen-z-in-an-uncertain-world/ [accessed 2023-10-09]
13. Singer K. Gen Zers Are Religious, And Their Mental Health Depends On It. Religion Unplugged. 2023. URL: https://religionunplugged.com/news/2023/1/6/gen-z-is-religious-and-their-mental-health-depends-on-it [accessed 2024-01-25]
14. Lucchetti G, Koenig HG, Lucchetti ALG. Spirituality, religiousness, and mental health: A review of the current scientific evidence. World J Clin Cases 2021 Sep 16;9(26):7620-7631 [FREE Full text] [doi: 10.12998/wjcc.v9.i26.7620] [Medline: 34621814]
15. Generation Z and Mental Health. The Annie E. Casey Foundation. URL: https://www.aecf.org/blog/generation-z-and-mental-health [accessed 2024-01-25]
16. Most U.S. Teens See Anxiety and Depression as a Major Problem Among Their Peers. Pew Research Center. URL: http://tinyurl.com/2rnxseh8 [accessed 2024-01-25]

17. Damen A, Exline J, Pargament K, Yao Y, Chochinov H, Emanuel L, et al. Prevalence, predictors and correlates of religious and spiritual struggles in palliative cancer patients. J Pain Symptom Manage 2021 Sep;62(3):e139-e147 [FREE Full text] [doi: 10.1016/j.jpainsymman.2021.04.024] [Medline: 33984462]
18. Barton KS, Tate T, Lau N, Taliesin KB, Waldman ED, Rosenberg AR. "I'm not a spiritual person." How hope might facilitate conversations about spirituality among teens and young adults with cancer. J Pain Symptom Manage 2018 Jun;55(6):1599-1608 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.02.001] [Medline: 29428188]
19. Spiritual Wellbeing. Australian National University. URL: https://www.anu.edu.au/spiritual-wellbeing [accessed 2024-01-25]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Covidence. URL: https://www.covidence.org/ [accessed 2024-01-25]
22. Park SY, Huberty J, Yourell J, McAlister KL, Beatty CC. A spiritual self-care mobile app (Skylight) for mental health, sleep, and spiritual well-being among generation Z and young millennials: Cross-sectional survey. JMIR Form Res 2023 Sep 27;7:e50239 [FREE Full text] [doi: 10.2196/50239] [Medline: 37597179]
23. Reed N, Peterson R, Ghost Dog T, Kaufman C, Kelley A, Craig Rushing S. Centering native youths' needs and priorities: Findings from the 2020 Native Youth Health Tech Survey. Am Indian Alsk Native Ment Health Res 2022 Sep;29(3):1-17. [doi: 10.5820/aian.2903.2022.1] [Medline: 36178745]
24. Au-Yeung A, Marfatia D, Beers K, General DA, McQueen KCD, Martin-Hill D, Six Nations Youth Mental Wellness Committee. Exploring the feasibility of a mental health application (JoyPop) for Indigenous youth. Front Psychiatry 2023 Oct 6;14:1269347 [FREE Full text] [doi: 10.3389/fpsyt.2023.1269347] [Medline: 37867769]
25. Mindu T, Mutero IT, Ngcobo WB, Musesengwa R, Chimbari MJ. Digital mental health interventions for young people in rural South Africa: Prospects and challenges for implementation. Int J Environ Res Public Health 2023 Jan 13;20(2):1453 [FREE Full text] [doi: 10.3390/ijerph20021453] [Medline: 36674209]
26. Gao C, Sun Y, Zhang F, Zhou F, Dong C, Ke Z, et al. Prevalence and correlates of lifestyle behavior, anxiety and depression in Chinese college freshman: A cross-sectional survey. Int J Nurs Sci 2021 Jul 10;8(3):347-353 [FREE Full text] [doi: 10.1016/j.ijnss.2021.05.013] [Medline: 34307785]
27. Spitzer R, Kroenke K, Williams J, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
28. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
29. Craig Rushing S, Kelley A, Bull S, Stephens D, Wrobel J, Silvasstar J, et al. Efficacy of an mHealth Intervention (BRAVE) to promote mental wellness for American Indian and Alaska native teenagers and young adults: Randomized controlled trial. JMIR Ment Health 2021 Sep 15;8(9):e26158 [FREE Full text] [doi: 10.2196/26158] [Medline: 34524092]
30. Hernandez BC, Vigna JF, Kelley ML. The Youth Coping Responses Inventory: development and initial validation. J Clin Psychol 2010 Sep;66(9):1008-1025. [doi: 10.1002/jclp.20697] [Medline: 20694963]
31. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: Systematic overview. JMIR Ment Health 2021 Apr 29;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
32. Gen Z mental health: The impact of tech and social media. McKinsey Health Institute. URL: https://www.mckinsey.com/mhi/our-insights/gen-z-mental-health-the-impact-of-tech-and-social-media [accessed 2024-01-25]
33. Health advisory on social media use in adolescence. APA. URL: https://www.apa.org/topics/social-media-internet/health-advisory-adolescent-social-media-use [accessed 2024-01-25]
34. When You Say You Believe In God, What Do You Mean? The Pew Charitable Trusts. URL: https://www.pewtrusts.org/en/trust/archive/fall-2018/when-you-say-you-believe-in-god-what-do-you-mean [accessed 2024-01-25]
35. Laird B, Puzia M, Larkey L, Ehlers D, Huberty J. A mobile app for stress management in middle-aged men and women (Calm): Feasibility randomized controlled trial. JMIR Form Res 2022 May 24;6(5):e30294 [FREE Full text] [doi: 10.2196/30294] [Medline: 34989677]
36. Neher T, Green J, Puzia M, Huberty J. Describing the use of a mindfulness-based app for sleep and mental well-being, across age, in children. Child Youth Care Forum 2021 Sep 24;51(4):749-768. [doi: 10.1007/s10566-021-09651-z]


## Abbreviations

**Gen Z:** Generation Z
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

*Edited by G Eysenbach, T de Azevedo Cardoso; submitted 11.05.23; peer-reviewed by F Dos Santos; comments to author 14.09.23; revised version received 11.10.23; accepted 15.01.24; published 06.02.24.*

*Please cite as:*
*Park SY, Do B, Yourell J, Hermer J, Huberty J*
*Digital Methods for the Spiritual and Mental Health of Generation Z: Scoping Review*
*Interact J Med Res 2024;13:e48929*
*URL: https://www.i-jmr.org/2024/1/e48929*
*doi:10.2196/48929*
*PMID:38261532*

Original Paper

# 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 Xu[1*], MPH; Jingya Zhou[2,3*], MPH; Hongxia Li[4], MD; Dong Cai[1], BA; Huanbing Zhu[5*], MD, PhD; Shengdong Pan[6*], MPA, LLM

[1]Medical Records Room, Department of Medical Administration, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2]Department of Medical Records, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[3]Collaborating Center for the WHO Family of International Classifications in China, Beijing, China

[4]Department of Gastroenterology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[5]Department of Hepatobiliary and Pancreatic Surgery, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[6]Department 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 ($\chi^2_1$=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 ($\chi^2_3$=93.7; *P*<.001), as well as a considerable difference between the changed and unchanged groups ($\chi^2_1$=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.

***(Interact J Med Res 2024;13:e52296)*** doi:10.2196/52296

**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.

## Methods

### Ethical Considerations

According to the Measures for Ethical Review of Human Life Science and Medical Research issued by the National Health Commission of the People's Republic of China, this study utilized public data and did not involve human subjects, and thus, the requirement of ethical permission was waived. All examples were constructed and neither correspond to real clinical cases nor to any datasets.

### 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 ϕ correlation coefficient was used to analyze the correlation between expression ability and grouping changes. Statistical significance was set at $P<.05$.

## Results

### Coding Features

#### *Coding Structure*

Chapter 2 in both the ICD-10 and ICD-11 addresses neoplasms, and the codes in both versions are alphanumeric but have different structures. The ICD-10 codes for neoplasms consist of topography codes and morphology codes. The topography codes range from C00 to D48. Except for leaf codes, the coding range, categories, and subcategories of the ICD-10-CCM are the same as those of the ICD-10. The ICD-10 morphology codes consist of 5 digits. The first 4 digits identify the histological type of the neoplasm, and the fifth digit, following a slash (/), indicates its behavior. In the ICD-10-CCM, the morphology codes consist of 6 digits, with a fifth number before the slash.

In contrast, the coding structure of the ICD-11 MMS codes has dramatically changed. It is composed of stem codes and extension codes that are connected by an ampersand (&). The precoordinated stem codes consist of sites and essential morphology types. The newly added chapter on extension codes addresses other morphology types and greater site specificity [11], as well as stage, grading, laterality, and the diagnostic method. These codes can be used for postcoordination. The stem codes range from 2A00 to 2F9Z. The extension codes for morphology are 6-digit codes composed of letters and numbers, starting with the letter *X*. Examples of the ICD-10 and ICD-11 complete neoplasm codes are shown in Table 1.

**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.

| Neoplasm | ICD-10 | | ICD-10-CCM | | ICD-11 | |
|---|---|---|---|---|---|---|
| | Topography code | Morphology code | Topography code | Morphology code | Stem code | Extension codes |
| Adenocarcinoma of the common bile duct, stage III, diagnosis confirmed by histology | C24.0 | 8140/3 | C24.003 | 81400/3 | 2C15.0 | • &XS6H<br>• &XY9Q |

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).

| Classification system | ICD-11 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 | Group 7 | Other chapters | No mapping |
| **ICD-10 codes, n** | | | | | | | | | |
| Group 1 | 23[a] | 91[a] | 352 | N/A[b] | N/A | N/A | N/A | 8[a] | 2 |
| Group 2 | N/A | N/A | N/A | 57 | N/A | N/A | N/A | N/A | N/A |
| Group 3 | 11[a] | N/A | 1[a] | N/A | 146 | N/A | N/A | 1[a] | N/A |
| Group 4 | 11[a] | 18[a] | N/A | N/A | N/A | 48[a] | 51[a] | 2[a] | 2 |
| **ICD-10-CCM codes, n** | | | | | | | | | |
| Group 1 | 93[a] | 499[a] | 1252 | N/A | N/A | N/A | N/A | 17[a] | 3 |
| Group 2 | N/A | N/A | N/A | 169 | N/A | N/A | N/A | N/A | N/A |
| Group 3 | 118[a] | N/A | 1[a] | N/A | 796 | N/A | N/A | 3[a] | N/A |
| Group 4 | 85[a] | 60[a] | N/A | N/A | N/A | 316[a] | 149[a] | 18[a] | 55 |

[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.

| ICD-10-CCM group | Total codes, N | Full representation | | | | Partial representation | |
|---|---|---|---|---|---|---|---|
| | | With PC[a] (a), n | Without PC (b), n | Codes (n1=a+b), n | Codes (n1/N), % | Codes (n2), n | Codes (n2/N), % |
| Group 1 | 1861 | 935 | 406 | 1341 | 72.1 | 520 | 27.9 |
| Group 2 | 169 | 106 | 44 | 150 | 88.8 | 19 | 11.2 |
| Group 3 | 918 | 501 | 73 | 574 | 62.5 | 344 | 37.5 |
| Group 4 | 628 | 288 | 71 | 359 | 57.2 | 269 | 42.8 |
| Total | 3576 | 1830 | 594 | 2424 | 67.8 | 1152 | 32.2 |

[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 ($\chi^2_1$=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 esophagogastric 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.

| Number and dimensions of postcoordination | Codes, n (%) |
|---|---|
| **0** | 1412 (39.5) |
| No | |
| **1** | |
| Specific anatomy | 1750 (48.9) |
| Histopathology | 176 (4.9) |
| Alternative severity 2 | 60 (1.7) |
| Laterality | 5 (0.1) |
| Course of the condition | 3 (0.1) |
| Stem codes | 3 (0.1) |
| **2** | |
| Specific anatomy/specific anatomy | 118 (3.3) |
| Specific anatomy/histopathology | 17 (0.5) |
| Specific anatomy/laterality | 8 (0.2) |
| Histopathology/alternative severity 2 | 8 (0.2) |
| Alternative severity 2/stem codes | 2 (0.1) |
| **3** | |
| Specific anatomy/specific anatomy/specific anatomy | 11 (0.3) |
| Specific anatomy/specific anatomy/laterality | 3 (0.1) |

## *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 ($\chi^2_1$=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 ($\chi^2_3$=93.7; *P*<.001), with ICD-10-CCM group 2 having the highest expression ability. The expression ability negatively correlated with grouping changes (*r*=–0.144; *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

1. ICD history. World Health Organization. 2022. URL: https://www.who.int/standards/classifications/classification-of-diseases [accessed 2022-11-16]
2. Peking Union Medical College Hospital, World Health Organization, International Classification Family Cooperation Center. The International Statistical Classification of Diseases and Related Health Problems tenth Revision (ICD-10)—volume 1-Tabular List. Beijing, China: People's Medical Publishing House; 2010.
3. Chute CG, Çelik C. Overview of ICD-11 architecture and structure. BMC Med Inform Decis Mak 2022;21(Suppl 6):378 [FREE Full text] [doi: 10.1186/s12911-021-01539-1] [Medline: 35578335]
4. Xu Y, Zhou J, Wang Y. The novel and powerful ICD-11 classification system for neoplasm coding: a comparative study with the ICD-O. BMC Med Inform Decis Mak 2022;22(1):333 [FREE Full text] [doi: 10.1186/s12911-022-02077-0] [Medline: 36527027]
5. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inform Decis Mak 2021;21(Suppl 6):206 [FREE Full text] [doi: 10.1186/s12911-021-01534-6] [Medline: 34753471]
6. National Health Commission of the People's Republic of China. China Health Statistical Yearbook. Beijing, China: NHC; 2023.
7. ICD-10/ICD-11 mapping tables. World Health Organization. 2022. URL: https://icd.who.int/browse11/l-m/en [accessed 2022-11-30]
8. Simple tabulation. World Health Organization. 2022. URL: https://icd.who.int/browse11/l-m/en [accessed 2022-11-30]
9. The second revision of the Chinese clinical modification of ICD-10 (ICD-10-CCM). National Health Commission of the People's Republic of China. 2019. URL: http://www.nhc.gov.cn/yzygj/s3593g/201904/b8323261bb8a4175a2046d2fffa93936.shtml [accessed 2022-04-15]
10. ICD-11 coding tool (version: 02/2022). World Health Organization. 2022. URL: https://icd.who.int/ct11/icd11_mms/zh/2022-02 [accessed 2024-01-13]
11. ICD-11 reference guide. World Health Organization. 2022. URL: https://icdcdn.who.int/icd11referenceguide/en/html/index.html [accessed 2022-06-26]
12. Fung KW, Xu J, Bodenreider O. The new international classification of diseases 11th edition: a comparative analysis with ICD-10 and ICD-10-CM. J Am Med Inform Assoc 2020;27(5):738-746 [FREE Full text] [doi: 10.1093/jamia/ocaa030] [Medline: 32364236]
13. Fung KW, Xu J, McConnell-Lamptey S, Pickett D, Bodenreider O. Feasibility of replacing the ICD-10-CM with the ICD-11 for morbidity coding: a content analysis. J Am Med Inform Assoc 2021;28(11):2404-2411 [FREE Full text] [doi: 10.1093/jamia/ocab156] [Medline: 34383897]

14. Aymé S, Bellet B, Rath A. Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. Orphanet J Rare Dis 2015;10:35 [FREE Full text] [doi: 10.1186/s13023-015-0251-8] [Medline: 25887186]
15. Chute CG. The rendering of human phenotype and rare diseases in ICD-11. J Inherit Metab Dis 2018;41(3):563-569 [FREE Full text] [doi: 10.1007/s10545-018-0172-5] [Medline: 29600497]
16. Zinboonyahgoon N, Luansritisakul C, Eiamtanasate S, Duangburong S, Sanansilp V, Korwisi B, et al. Comparing the ICD-11 chronic pain classification with ICD-10: how can the new coding system make chronic pain visible? A study in a tertiary care pain clinic setting. Pain 2021;162(7):1995-2001. [doi: 10.1097/j.pain.0000000000002196] [Medline: 33449507]
17. Mabon K, Steinum O, Chute CG. Postcoordination of codes in ICD-11. BMC Med Inform Decis Mak 2022;21(Suppl 6):379 [FREE Full text] [doi: 10.1186/s12911-022-01876-9] [Medline: 35581649]
18. Drösler SE, Weber S, Chute CG. ICD-11 extension codes support detailed clinical abstraction and comprehensive classification. BMC Med Inform Decis Mak 2021;21(Suppl 6):278 [FREE Full text] [doi: 10.1186/s12911-021-01635-2] [Medline: 34753461]
19. Tavakoli K, Kalaw FGP, Bhanvadia S, Hogarth M, Baxter SL. Concept coverage analysis of ophthalmic infections and trauma among the standardized medical terminologies SNOMED-CT, ICD-10-CM, and ICD-11. Ophthalmol Sci 2023;3(4):100337 [FREE Full text] [doi: 10.1016/j.xops.2023.100337] [Medline: 37449050]
20. Rodrigues JM, Schulz S, Rector A, Spackman K, Millar J, Campbell J, et al. ICD-11 and SNOMED CT common ontology: circulatory system. Stud Health Technol Inform 2014;205:1043-1047. [Medline: 25160347]
21. Rodrigues JM, Schulz S, Rector A, Spackman K, Üstün B, Chute CG, et al. Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability. Stud Health Technol Inform 2013;192:343-346. [Medline: 23920573]
22. Rodrigues JM, Robinson D, Della Mea V, Campbell J, Rector A, Schulz S, et al. Semantic alignment between ICD-11 and SNOMED CT. Stud Health Technol Inform 2015;216:790-794. [Medline: 26262160]
23. Zarei J, Golpira R, Hashemi N, Azadmanjir Z, Meidani Z, Vahedi A, et al. Comparison of the accuracy of inpatient morbidity coding with ICD-11 and ICD-10. Health Inf Manag 2023;0(0):18333583231185355. [doi: 10.1177/18333583231185355] [Medline: 37491819]
24. Lee H, Kim S. Impact of the ICD-11 on the accuracy of clinical coding in Korea. Health Inf Manag 2023;52(3):221-228. [doi: 10.1177/18333583221095147] [Medline: 35614871]
25. Eisele A, Dereskewitz C, Oberhauser C, Kus S, Coenen M. Reliability, usability and applicability of the ICD-11 beta draft focusing on hand injuries and diseases: results from German field testing. Int J Qual Health Care 2019;31(10):G174-G179 [FREE Full text] [doi: 10.1093/intqhc/mzz121] [Medline: 31838492]
26. Eastwood CA, Southern DA, Doktorchik C, Khair S, Cullen D, Boxill A, et al. Training and experience of coding with the World Health Organization's international classification of diseases, eleventh revision. Health Inf Manag 2023;52(2):92-100 [FREE Full text] [doi: 10.1177/18333583211038633] [Medline: 34555947]
27. Monestime JP, Mayer RW, Blackwood A. Analyzing the ICD-10-CM transition and post-implementation stages: a public health institution case study. Perspect Health Inf Manag 2019;16(Spring):1a [FREE Full text] [Medline: 31019430]
28. Libicki M, Brahmakulam I. The costs and benefits of moving to the ICD-10 code sets. RAND Corporation. 2004. URL: https://www.rand.org/pubs/technical_reports/TR132.html [accessed 2024-01-13]
29. Kusnoor SV, Blasingame MN, Williams AM, DesAutels SJ, Su J, Giuse NB. A narrative review of the impact of the transition to ICD-10 and ICD-10-CM/PCS. JAMIA Open 2020;3(1):126-131 [FREE Full text] [doi: 10.1093/jamiaopen/ooz066] [Medline: 32607494]
30. Xu J, Fung KW, Bodenreider O. Sequential mapping—a novel approach to map from ICD-10-CM to ICD-11. Stud Health Technol Inform 2022;290:96-100 [FREE Full text] [doi: 10.3233/SHTI220039] [Medline: 35672978]
31. Hong N, Zhang Y, Ren Y, Hou L, Wang C, Li J, et al. Cross-language terminology mapping between ICD-10-CN and SNOMED-CT. Stud Health Technol Inform 2022;290:42-46. [doi: 10.3233/SHTI220028] [Medline: 35672967]
32. Ohlsen T, Kruse V, Krupar R, Banach A, Ingenerf J, Drenkhahn C. Mapping of ICD-O tuples to OncoTree codes using SNOMED CT post-coordination. Stud Health Technol Inform 2022;294:307-311. [doi: 10.3233/SHTI220464] [Medline: 35612082]
33. Grosjean J, Billey K, Charlet J, Darmoni SJ. Manual evaluation of the automatic mapping of International Classification of Diseases (ICD)-11 (in French). Stud Health Technol Inform 2020;270:1335-1336. [doi: 10.3233/SHTI200429] [Medline: 32570646]
34. Caskey RN, Abutahoun A, Polick A, Barnes M, Srivastava P, Boyd AD. Transition to international classification of disease version 10, clinical modification: the impact on internal medicine and internal medicine subspecialties. BMC Health Serv Res 2018;18(1):328 [FREE Full text] [doi: 10.1186/s12913-018-3110-1] [Medline: 29728145]
35. Sun D, Li H, Cao M, He S, Lei L, Peng J, et al. Cancer burden in China: trends, risk factors and prevention. Cancer Biol Med 2020;17(4):879-895 [FREE Full text] [doi: 10.20892/j.issn.2095-3941.2020.0387] [Medline: 33299641]

## 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

*Edited by M Sheth; submitted 29.08.23; peer-reviewed by T Karen, D Gartner; comments to author 28.11.23; revised version received 13.01.24; accepted 31.01.24; published 08.03.24.*

*Please cite as:*
*Xu Y, Zhou J, Li H, Cai D, Zhu H, Pan S*
*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*
*Interact J Med Res 2024;13:e52296*
*URL: https://www.i-jmr.org/2024/1/e52296*
*doi:10.2196/52296*
*PMID:38457228*

Original

# Application of AI in Sepsis: Citation Network Analysis and Evidence Synthesis

MeiJung Wu[1,2], MSc; Md Mohaimenul Islam[3], PhD; Tahmina Nasrin Poly[1], MSc; Ming-Chin Lin[1,4,5], PhD

[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[2]Department of Nursing, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[3]Department of Outcomes and Translational Sciences, College of Pharmacy, The Ohio State University, Columbus, OH, United States

[4]Taipei Neuroscience Institute, Taipei Medical University, Taipei, Taiwan

[5]Department 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.

***(Interact J Med Res 2024;13:e54490)*** doi:10.2196/54490

**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 bibliometric 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:

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.

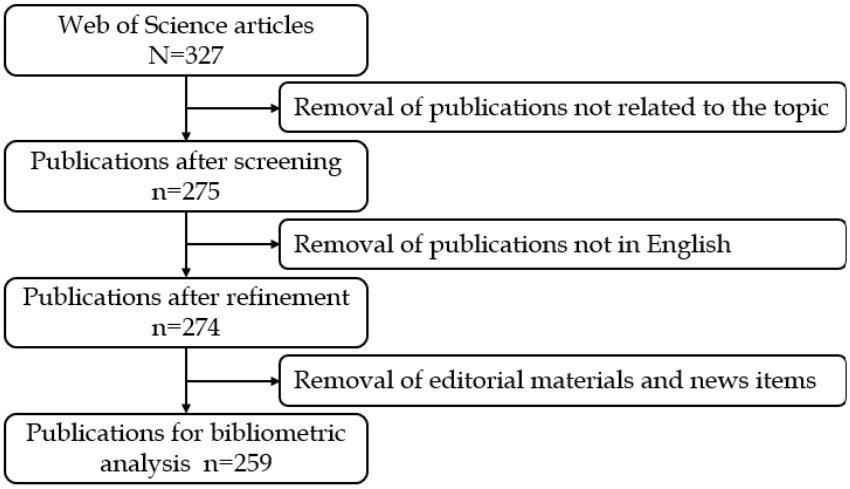
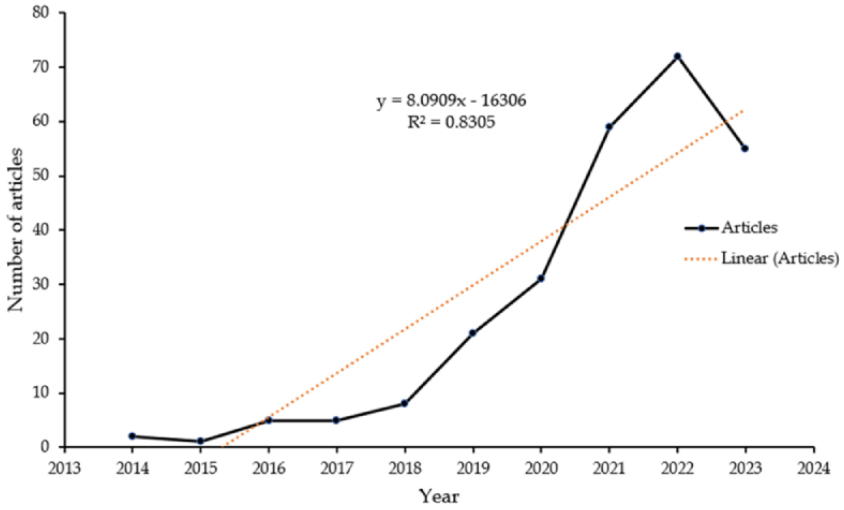



**Figure 2.** Trends in the number of publications on the application of artificial intelligence to the study of sepsis from 2014 to 2023 (August).




### 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.

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

## 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.

| Rank | Country | Articles, n (%) |
|---|---|---|
| 1 | United States | 103 (39.8) |
| 2 | China | 83 (32) |
| 3 | United Kingdom | 14 (5.4) |
| 4 | Taiwan | 12 (4.6) |
| 5 | India | 11 (4.2) |
| 6 | Netherlands | 10 (3.9) |
| 7 | Australia | 8 (3.1) |
| 8 | Canada | 8 (3.1) |
| 9 | Spain | 7 (2.7) |
| 10 | Germany | 7 (2.7) |

## Distribution of Institutions

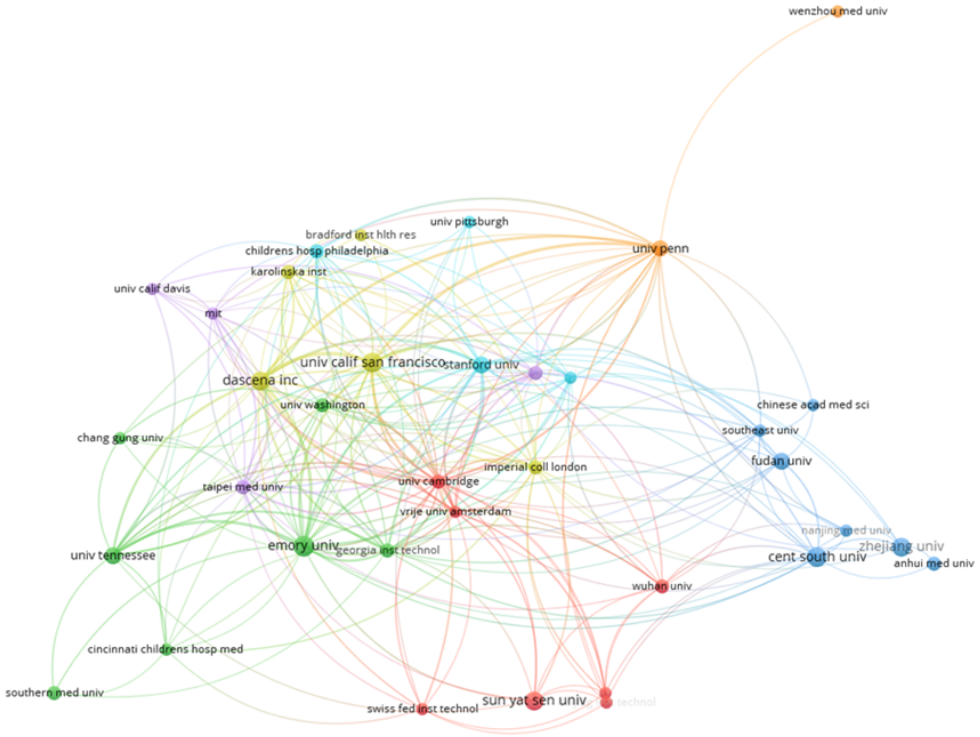
Table 3 shows the top 10 most productive institutes that used AI applications in sepsis. The University of California system (22/259 articles, 8.5%) ranked first among all research institutions, followed by Emory University (10/259 articles, 3.9%), Harvard University (10/259 articles, 3.9%), and Central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

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

**Figure 3.** Institution co-operation network.




## Distribution of Authors

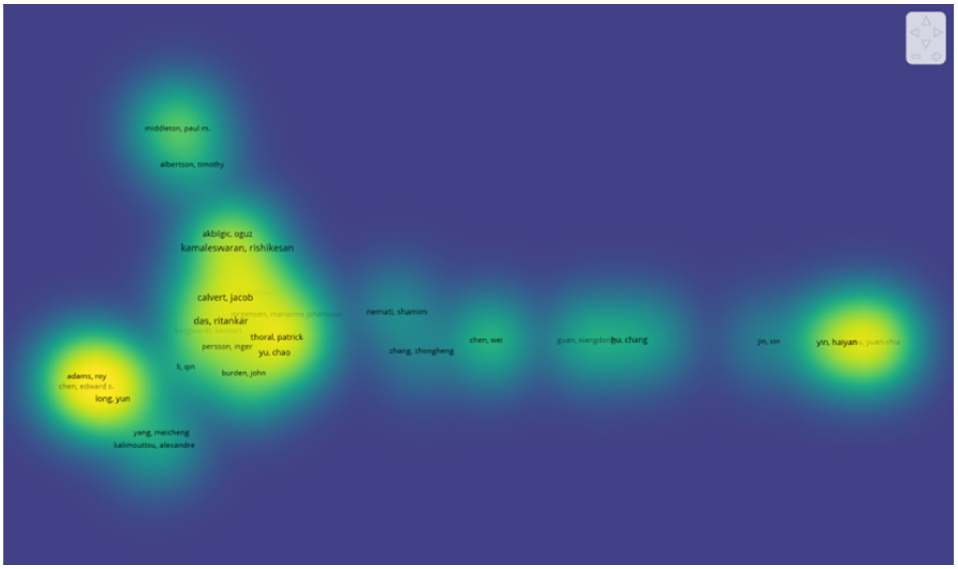
A total of 259 articles were authored by 1444 individuals, each of whom had at least 1 article to their name. In Table 4, we present the top 10 most prolific authors who conducted and published research in these fields. Ritankar Das secured the top position with 8 articles, closely followed by Chris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

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

**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.

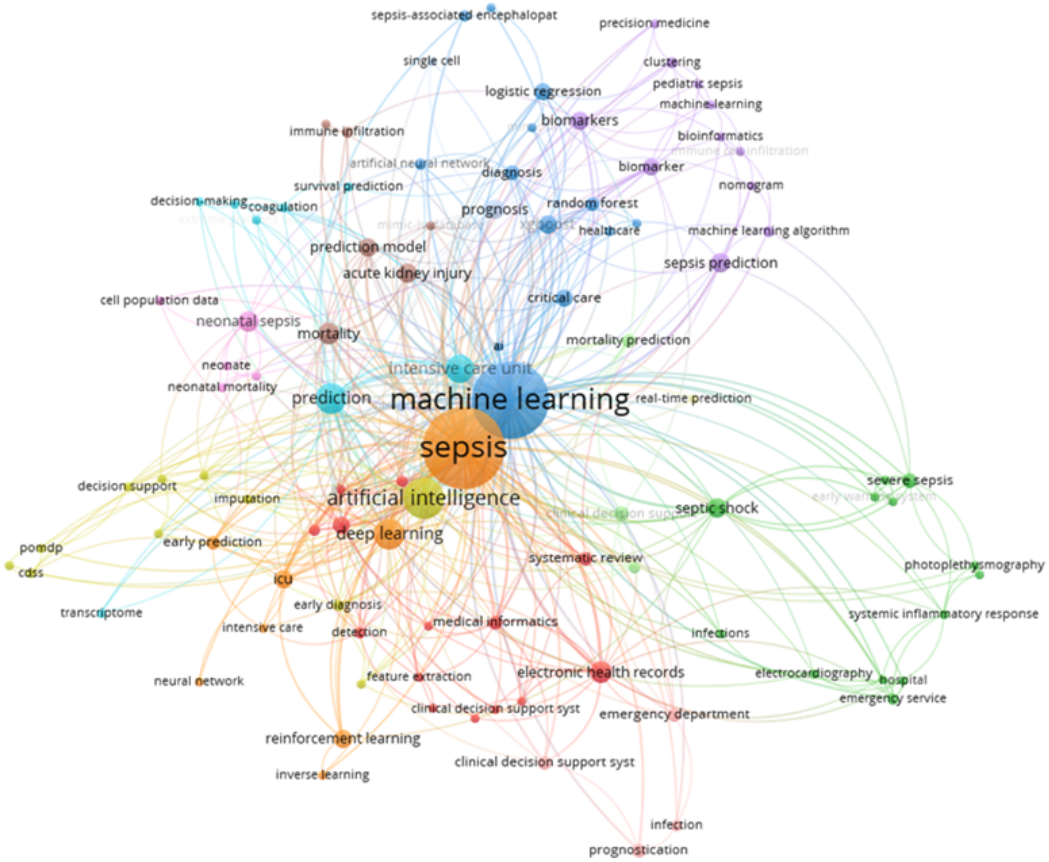
| Rank | Author | Journal | Title | Citation, n |
|---|---|---|---|---|
| 1 | Komorowski et al [16] | *Nature Medicine* | The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care | 408 |
| 2 | Nemati et al 2018 [17] | *Critical Care Medicine* | An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU | 329 |
| 3 | Taylor et al 2016 [18] | *Academic Emergency Medicine* | Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach | 257 |
| 4 | Desautels et al 2016 [19] | *JMIR Medical Informatics* | Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach | 226 |
| 5 | Fleuren et al [10] | *Intensive Care Medicine* | Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy | 184 |
| 6 | Horng et al 2017 [20] | *PlosOne* | Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning | 148 |
| 7 | Gultepe et al 2014 [21] | *Journal of the American Medical Informatics Association* | From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system | 101 |
| 8 | Giannini et al 2019 [22] | *Critical Care Medicine* | A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice | 100 |
| 9 | Hou et al 2020 [23] | *Journal of Translational Medicine* | Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost | 98 |
| 10 | Mani et al 2014 [24] | *Journal of the American Medical Informatics Association* | Medical decision support using machine learning for early detection of late-onset neonatal sepsis | 97 |

## 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).

**Figure 5.** The co-occurrence network of the top 100 keywords in artificial intelligence research for sepsis from 2014 to 2023 (August).




## 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

1. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed 2019;170:1-9. [doi: 10.1016/j.cmpb.2018.12.027] [Medline: 30712598]
2. Wang W, Liu CF. Sepsis heterogeneity. World J Pediatr 2023;19(10):919-927. [doi: 10.1007/s12519-023-00689-8] [Medline: 36735197]
3. Wardi G, Tainter CR, Ramnath VR, Brennan JJ, Tolia V, Castillo EM, et al. Age-related incidence and outcomes of sepsis in California, 2008-2015. J Crit Care 2021;62:212-217 [FREE Full text] [doi: 10.1016/j.jcrc.2020.12.015] [Medline: 33429114]
4. Rhee C, Jones TM, Hamad Y, Pande A, Varon J, O'Brien C, et al. Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. JAMA Netw Open 2019;2(2):e187571 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.7571] [Medline: 30768188]
5. Vincent JL. The clinical challenge of sepsis identification and monitoring. PLoS Med 2016;13(5):e1002022 [FREE Full text] [doi: 10.1371/journal.pmed.1002022] [Medline: 27187803]
6. Talisa VB, Yende S, Seymour CW, Angus DC. Arguing for adaptive clinical trials in sepsis. Front Immunol 2018;9:1502 [FREE Full text] [doi: 10.3389/fimmu.2018.01502] [Medline: 30002660]
7. de Grooth HJ, Postema J, Loer SA, Parienti JJ, Oudemans-van Straaten HM, Girbes AR. Unexplained mortality differences between septic shock trials: a systematic analysis of population characteristics and control-group mortality rates. Intensive Care Med 2018;44(3):311-322 [FREE Full text] [doi: 10.1007/s00134-018-5134-8] [Medline: 29546535]
8. Goh KH, Wang L, Yeow AYK, Poh H, Li K, Yeow JJL, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nat Commun 2021;12(1):711 [FREE Full text] [doi: 10.1038/s41467-021-20910-4] [Medline: 33514699]
9. Luz CF, Vollmer M, Decruyenaere J, Nijsten MW, Glasner C, Sinha B. Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. Clin Microbiol Infect 2020;26(10):1291-1299 [FREE Full text] [doi: 10.1016/j.cmi.2020.02.003] [Medline: 32061798]
10. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020;46(3):383-400 [FREE Full text] [doi: 10.1007/s00134-019-05872-y] [Medline: 31965266]

11. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res 2020;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]
12. Chen X, Xie H, Li Z, Cheng G, Leng M, Wang FL. Information fusion and artificial intelligence for smart healthcare: a bibliometric study. Inf Process Manage 2023;60(1):103113 [FREE Full text] [doi: 10.1016/j.ipm.2022.103113]
13. Wu CC, Huang CW, Wang YC, Islam MM, Kung WM, Weng YC, et al. mHealth research for weight loss, physical activity, and sedentary behavior: bibliometric analysis. J Med Internet Res 2022;24(6):e35747 [FREE Full text] [doi: 10.2196/35747] [Medline: 35675126]
14. Poly TN, Islam MM, Walther BA, Lin MC, Jack Li YC. Artificial intelligence in diabetic retinopathy: bibliometric analysis. Comput Methods Programs Biomed 2023;231:107358. [doi: 10.1016/j.cmpb.2023.107358] [Medline: 36731310]
15. Yeung AWK, Parvanov ED, Hribersek M, Eibensteiner F, Klager E, Kletecka-Pulker M, et al. Digital teaching in medical education: scientific literature landscape review. JMIR Med Educ 2022;8(1):e32747 [FREE Full text] [doi: 10.2196/32747] [Medline: 35138260]
16. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018;24(11):1716-1720 [FREE Full text] [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
17. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med 2018;46(4):547-553 [FREE Full text] [doi: 10.1097/CCM.0000000000002936] [Medline: 29286945]
18. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. Acad Emerg Med 2016;23(3):269-278 [FREE Full text] [doi: 10.1111/acem.12876] [Medline: 26679719]
19. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inform 2016;4(3):e28 [FREE Full text] [doi: 10.2196/medinform.5909] [Medline: 27694098]
20. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One 2017;12(4):e0174708 [FREE Full text] [doi: 10.1371/journal.pone.0174708] [Medline: 28384212]
21. Gultepe E, Green JP, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. J Am Med Inform Assoc 2014;21(2):315-325 [FREE Full text] [doi: 10.1136/amiajnl-2013-001815] [Medline: 23959843]
22. Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. Crit Care Med 2019;47(11):1485-1492 [FREE Full text] [doi: 10.1097/CCM.0000000000003891] [Medline: 31389839]
23. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. J Transl Med 2020;18(1):462 [FREE Full text] [doi: 10.1186/s12967-020-02620-5] [Medline: 33287854]
24. Mani S, Ozdas A, Aliferis C, Varol HA, Chen Q, Carnevale R, et al. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. J Am Med Inform Assoc 2014;21(2):326-336 [FREE Full text] [doi: 10.1136/amiajnl-2013-001854] [Medline: 24043317]
25. Jimma BL. Artificial intelligence in healthcare: a bibliometric analysis. Telemat Inform Rep 2023;9:100041 [FREE Full text] [doi: 10.1016/j.teler.2023.100041]
26. Hong C, Wang L. Virtual reality technology in nursing professional skills training: bibliometric analysis. JMIR Serious Games 2023;11:e44766 [FREE Full text] [doi: 10.2196/44766] [Medline: 37603389]
27. Tang R, Zhang S, Ding C, Zhu M, Gao Y. Artificial intelligence in intensive care medicine: bibliometric analysis. J Med Internet Res 2022;24(11):e42185 [FREE Full text] [doi: 10.2196/42185] [Medline: 36449345]
28. Wang J, Liang Y, Cao S, Cai P, Fan Y. Application of artificial intelligence in geriatric care: bibliometric analysis. J Med Internet Res 2023;25:e46014 [FREE Full text] [doi: 10.2196/46014] [Medline: 37351923]
29. Suhail F, Adel M, Al-Emran M, Shaalan K. A bibliometric analysis on the role of artificial intelligence in healthcare. In: Mishra S, Tripathy HK, Mallick P, Shaalan K, editors. Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis. Singapore: Springer; 2022:1-14.
30. Javidan AP, Li A, Lee MH, Forbes TL, Naji F. A systematic review and bibliometric analysis of applications of artificial intelligence and machine learning in vascular surgery. Ann Vasc Surg 2022;85:395-405. [doi: 10.1016/j.avsg.2022.03.019] [Medline: 35339595]
31. Gundersen T, Bærøe K. The future ethics of artificial intelligence in medicine: making sense of collaborative models. Sci Eng Ethics 2022;28(2):17 [FREE Full text] [doi: 10.1007/s11948-022-00369-2] [Medline: 35362822]
32. Rowley J, Sbaffi L, Sugden M, Gilbert A. Factors influencing researchers' journal selection decisions. J Inf Sci 2022;48(3):321-335 [FREE Full text] [doi: 10.1177/0165551520958591]
33. Triggle CR, MacDonald R, Triggle DJ, Grierson D. Requiem for impact factors and high publication charges. Account Res 2022;29(3):133-164 [FREE Full text] [doi: 10.1080/08989621.2021.1909481] [Medline: 33787413]

34. Hardwicke TE, Thibault RT, Kosie JE, Tzavella L, Bendixen T, Handcock SA, et al. Post-publication critique at top-ranked journals across scientific disciplines: a cross-sectional assessment of policies and practice. R Soc Open Sci 2022;9(8):220139 [FREE Full text] [doi: 10.1098/rsos.220139] [Medline: 36039285]
35. Bartocci L, Grossi G, Mauro SG, Ebdon C. The journey of participatory budgeting: a systematic literature review and future research directions. Int Rev Adm Sci 2023;89(3):757-774 [FREE Full text] [doi: 10.1177/00208523221078938]
36. Maddi A, Sapinho D. Article processing charges, altmetrics and citation impact: is there an economic rationale? Scientometrics 2022;127(12):7351-7368. [doi: 10.1007/s11192-022-04284-y]
37. Yan MY, Gustad LT, Nytrø Ø. Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. J Am Med Inform Assoc 2022;29(3):559-575 [FREE Full text] [doi: 10.1093/jamia/ocab236] [Medline: 34897469]
38. Deng HF, Sun MW, Wang Y, Zeng J, Yuan T, Li T, et al. Evaluating machine learning models for sepsis prediction: a systematic review of methodologies. iScience 2022;25(1):103651 [FREE Full text] [doi: 10.1016/j.isci.2021.103651] [Medline: 35028534]
39. Kijpaisalratana N, Sanglertsinlapachai D, Techaratsami S, Musikatavorn K, Saoraya J. Machine learning algorithms for early sepsis detection in the emergency department: a retrospective study. Int J Med Inform 2022;160:104689. [doi: 10.1016/j.ijmedinf.2022.104689] [Medline: 35078027]
40. Yeung AWK, Kulnik ST, Parvanov ED, Fassl A, Eibensteiner F, Völkl-Kernstock S, et al. Research on digital technology use in cardiology: bibliometric analysis. J Med Internet Res 2022;24(5):e36086 [FREE Full text] [doi: 10.2196/36086] [Medline: 35544307]
41. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022;24(7):e35816 [FREE Full text] [doi: 10.2196/35816] [Medline: 35793141]
42. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Tsagkaris C, Parvanov ED, et al. Medical and health-related misinformation on social media: bibliometric study of the scientific literature. J Med Internet Res 2022;24(1):e28152 [FREE Full text] [doi: 10.2196/28152] [Medline: 34951864]


## Abbreviations

**AI:** artificial intelligence
**MIMIC-III:** Multiparameter Intelligent Monitoring in Intensive Care III


*Edited by T de Azevedo Cardoso; submitted 11.11.23; peer-reviewed by E Kawamoto, D Amancio; comments to author 19.01.24; revised version received 27.01.24; accepted 04.03.24; published 15.04.24.*

*Please cite as:*
*Wu M, Islam MM, Poly TN, Lin MC*
*Application of AI in Sepsis: Citation Network Analysis and Evidence Synthesis*
*Interact J Med Res 2024;13:e54490*
*URL: https://www.i-jmr.org/2024/1/e54490*
*doi:10.2196/54490*
*PMID:38621231*

Review

# Theoretical Perspectives Underpinning Research on the Physician-Patient Relationship in a Digital Health Practice: Scoping Review

Damoun Nassehi[1], MD, PhD; Birgitta Haga Gripsrud[1], Prof Dr; Ellen Ramvi[1], Prof Dr

Department of Caring and Ethics, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway

**Corresponding Author:**
Damoun Nassehi, MD, PhD
Department of Caring and Ethics
Faculty of Health Sciences
University of Stavanger
Kjell Arholms gate 41
Stavanger, 4036
Norway
Phone: 47 45426096
Email: damoun.nassehi@uis.no

## Abstract

**Background:** 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.

**Objective:** 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.

**Methods:** 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.

**Results:** 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.

**Conclusions:** 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.

***(Interact J Med Res 2024;13:e47280)*** doi:10.2196/47280

**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.

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

[a]mHealth: 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.

| Authors | Year | Journal | Country | Study type | Type of technology | Aims | Participants |
|---|---|---|---|---|---|---|---|
| Audrain-Pontevia and Menvielle [15] | 2018 | *Health Services Management Research* | Canada | Quantitative | Online health community | Examine how online health communities impact the physician-patient relationship | Patients |
| Balato et al [16] | 2013 | *British Journal of Dermatology* | Italy | Quantitative | Mobile phone messages | Evaluate the use of telemedicine in improving treatment adherence, patient outcomes, and the physician-patient relationship | Patients and physicians |
| Grünloh et al [17] | 2018 | *Journal of Medical Internet Research* | Sweden | Qualitative | Web-based patient portal | Investigate how physicians view the idea of patient participation | Physicians |
| Győrffy et al [18] | 2020 | *PLOS ONE* | Hungary | Qualitative | Social media | Explore physicians' knowledge and attitudes toward digital health technologies and the transformation of the doctor-patient relationship | Physicians |
| Jiang [19] | 2019 | *Health Communication* | China | Quantitative | Digital messaging | Examine how the quality of face-to-face communication with providers is associated with their subsequent internet use for patient-provider communication | Patients |
| Kludacz-Alessandri et al [20] | 2021 | *PLOS ONE* | Poland | Qualitative | Teleconsultation (phone) | Study patients' satisfaction with teleconsultation in primary care and the impact of teleconsultations on GP[a]-patient communication | Patients and physicians |
| Macdonald et al [21] | 2018 | *Journal of Medical Internet Research* | Canada | Qualitative | Digital messaging and EHR[b] | Examine HCP's[c] perspectives on how eHealth affects their relationships with patients, as well as its ethical ramifications | Patients, physicians, and other HCPs |
| Tasneem et al [22] | 2019 | *American Journal of Hospice and Palliative Medicine* | United States | Qualitative | Video consultation | Investigate the need for web-based videoconferences for oncology patients | Patients and physicians |
| Townsend et al [23] | 2015 | *Journal of Medical Internet Research* | Canada | Qualitative | Health-related websites | Focus on patients' and HCP's use of health-related internet information and how it influences the patient-HCP relationship | Patients, physicians, and other HCPs |
| Yan et al [24] | 2020 | *International Journal of Environmental Research and Public Health* | China | Qualitative | Digital messaging | Understand the underlying reasons for poor doctor-patient relationships | Patients and physicians |

[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).

**Table 3.** Summary of evidence.

| Authors | Summary: What does the paper state about the physician-patient relationship? |
|---|---|
| Audrain-Pontevia and Menvielle [15] | • 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.<br>• 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.<br>• 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. |
| Balato et al [16] | • 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. |
| Grünloh et al [17] | • 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.<br>• The paper suggests that the use of patient-accessible EHRs[a] 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.<br>• 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.<br>• The authors mention the importance of patient participation for patient safety. |
| Győrffy et al [18] | • 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.<br>• 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.<br>• 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.<br>• Medical school curricula should emphasize health and prevention rather than only diseases and pathology via the newest digital technological solutions.<br>• 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. |
| Jiang [19] | • The results of the study emphasize the important roles of patient-centered communication and the physician-patient relationship in the eHealth and mHealth[b] movement, particularly in the Chinese health care system.<br>• 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.<br>• 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). |
| Kludacz-Alessandri et al [20] | • The study concerns patients' satisfaction with teleconsultation in primary care and the impact of teleconsultations on GP[c]-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. |
| Macdonald et al [21] | • 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.<br>• 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.<br>• One of the health care professionals stated that patients who are engaged through eHealth and informed about their condition are more useful clinically.<br>• The paper also mentions a pedagogical approach to interacting with the eHealth users among patients. |

| Authors | Summary: What does the paper state about the physician-patient relationship? |
|---|---|
| Tasneem et al [22] | • 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. |
| Townsend et al [23] | • 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.<br>• 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.<br>• 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. |
| Yan et al [24] | • 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.<br>• 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.<br>• 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. |

[a]EHR: electronic health record.

[b]mHealth: mobile health.

[c]GP: general practitioner.

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

| Authors | Framework or concepts | Relevance of using specific theoretical frameworks and concepts |
|---|---|---|
| Audrain-Pontevia and Menvielle [15] | Patient empowerment | Evaluates the impact of online health communities on the physician-patient relationship. |
| Balato et al [16] | None mentioned | N/A[a] |
| Grünloh et al [17] | Shared decision-making, patient-centered care, and paternalism | Assesses the impact of a web-based patient portal on patient participation. |
| Győrffy et al [18] | Patient empowerment and shared decision-making | Studies the attitudes of digitally engaged physicians toward transforming the physician-patient relationship. |
| Jiang [19] | Patient-centered care, mutual trust, and patient satisfaction | Examines the relationship between face-to-face and online patient-provider communication. |
| Kludacz-Alessandri et al [20] | Mutual trust | Investigates the satisfaction of patients with teleconsultations. |
| Macdonald et al [21] | Relational ethics | Relational ethics addresses the ethical content and decisions implicit in everyday relationships and conversations. |
| Tasneem et al [22] | None mentioned | N/A |
| Townsend et al [23] | Relational ethics | 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. |
| Yan et al [24] | None mentioned | N/A |

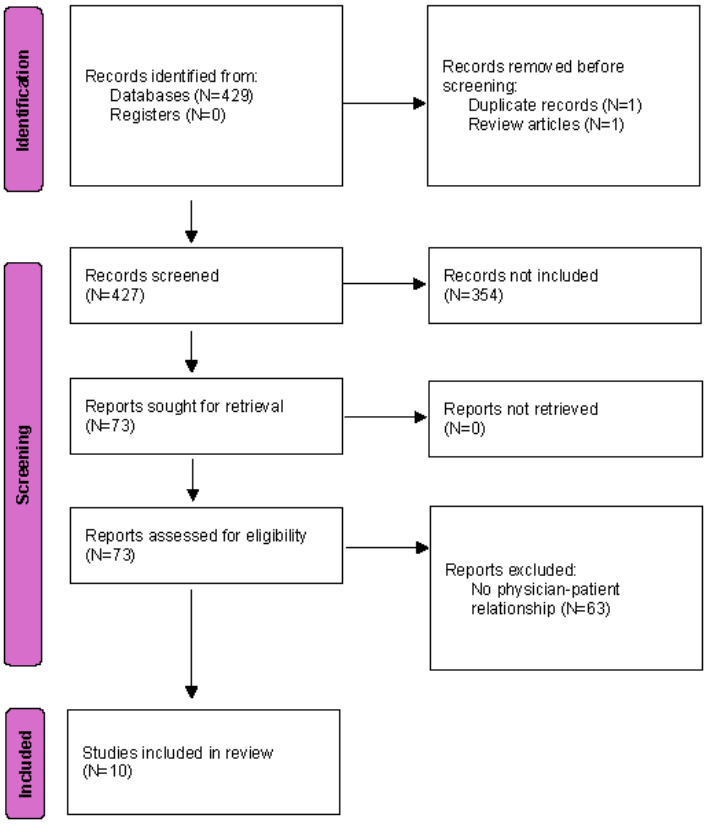
[a]N/A: not applicable.

# 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. 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 Grünloh 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.

Győrffy 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 - ijmr_v13i1e47280_app1.pdf ]

## References

1. Mesko B. Digital health technologies and well-being in the future. IT Prof 2020 Jan 1;22(1):20-23 [FREE Full text] [doi: 10.1109/mitp.2019.2963121]
2. Wadhwa V. Exponential health technology bringing personal 'check engine lights'. Forbes. 2013. URL: https://www.forbes.com/sites/singularity/ [accessed 2021-03-26]
3. Vogt H, Ulvestad E, Eriksen TE, Getz L. Getting personal: can systems medicine integrate scientific and humanistic conceptions of the patient? J Eval Clin Pract 2014;20(6):942-952 [FREE Full text] [doi: 10.1111/jep.12251] [Medline: 25312489]
4. Rudebeck CE. Relationship based care - how general practice developed and why it is undermined within contemporary healthcare systems. Scand J Prim Health Care 2019;37(3):335-344 [FREE Full text] [doi: 10.1080/02813432.2019.1639909] [Medline: 31299870]
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56 [FREE Full text] [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
6. Boers SN, Jongsma KR, Lucivero F, Aardoom J, Büchner FL, de Vries M, et al. SERIES: eHealth in primary care. part 2: exploring the ethical implications of its application in primary care practice. Eur J Gen Pract 2020;26(1):26-32 [FREE Full text] [doi: 10.1080/13814788.2019.1678958] [Medline: 31663394]
7. Sandvik H, Hetlevik Ø, Blinkenberg J, Hunskaar S. Continuity in general practice as predictor of mortality, acute hospitalisation, and use of out-of-hours care: a registry-based observational study in Norway. Br J Gen Pract 2022;72(715):e84-e90 [FREE Full text] [doi: 10.3399/BJGP.2021.0340] [Medline: 34607797]
8. Björndell C, Premberg Å. Physicians' experiences of video consultation with patients at a public virtual primary care clinic: a qualitative interview study. Scand J Prim Health Care 2021;39(1):67-76 [FREE Full text] [doi: 10.1080/02813432.2021.1882082] [Medline: 33650941]
9. The Lancet Digital Health. A digital (r)evolution: introducing the lancet digital health. Lancet Digit Health 2019;1(1):e1 [FREE Full text] [doi: 10.1016/S2589-7500(19)30010-X] [Medline: 33323234]
10. Moore J, Engel J, Prentice D. Relational ethics in everyday practice. Can Oncol Nurs J 2014;24(1):31-39. [doi: 10.5737/1181912x2413134] [Medline: 24707705]
11. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
12. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
13. Nassehi D. Physicians’ use of digital technologies and its impact on the physician-patient relationship - a scoping review. OSF Registries. 2022. URL: https://osf.io/jkq2f [accessed 2023-03-12]
14. Bock W, Candelon F, Chai S, Choi E, Corwin J, DiGrande S, et al. The mobile revolution: how mobile technologies drive a trillion-dollar impact. The Boston Consulting Group. 2015. URL: https://www.bcg.com/publications/2015/telecommunications-technology-industries-the-mobile-revolution [accessed 2023-11-25]
15. Audrain-Pontevia AF, Menvielle L. Do online health communities enhance patient-physician relationship? an assessment of the impact of social support and patient empowerment. Health Serv Manage Res 2018;31(3):154-162 [FREE Full text] [doi: 10.1177/0951484817748462] [Medline: 29280679]

16. Balato N, Megna M, Di Costanzo L, Balato A, Ayala F. Educational and motivational support service: a pilot study for mobile-phone-based interventions in patients with psoriasis. Br J Dermatol 2013;168(1):201-205 [FREE Full text] [doi: 10.1111/j.1365-2133.2012.11205.x] [Medline: 23240729]
17. Grünloh C, Myreteg G, Cajander Å, Rexhepi H. "Why do they need to check me?" patient participation through ehealth and the doctor-patient relationship: qualitative study. J Med Internet Res 2018;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
18. Győrffy Z, Radó N, Mesko B. Digitally engaged physicians about the digital health transition. PLoS One 2020;15(9):e0238658 [FREE Full text] [doi: 10.1371/journal.pone.0238658] [Medline: 32986733]
19. Jiang S. The relationship between face-to-face and online patient-provider communication: examining the moderating roles of patient trust and patient satisfaction. Health Commun 2020;35(3):341-349 [FREE Full text] [doi: 10.1080/10410236.2018.1563030] [Medline: 30821506]
20. Kludacz-Alessandri M, Hawrysz L, Korneta P, Gierszewska G, Pomaranik W, Walczak R. The impact of medical teleconsultations on general practitioner-patient communication during COVID- 19: a case study from Poland. PLoS One 2021;16(7):e0254960 [FREE Full text] [doi: 10.1371/journal.pone.0254960] [Medline: 34270587]
21. Macdonald GG, Townsend AF, Adam P, Li LC, Kerr S, McDonald M, et al. eHealth technologies, multimorbidity, and the office visit: qualitative interview study on the perspectives of physicians and nurses. J Med Internet Res 2018;20(1):e31 [FREE Full text] [doi: 10.2196/jmir.8983] [Medline: 29374004]
22. Tasneem S, Kim A, Bagheri A, Lebret J. Telemedicine video visits for patients receiving palliative care: a qualitative study. Am J Hosp Palliat Care 2019;36(9):789-794 [FREE Full text] [doi: 10.1177/1049909119846843] [Medline: 31064195]
23. Townsend A, Leese J, Adam P, McDonald M, Li LC, Kerr S, et al. eHealth, participatory medicine, and ethical care: a focus group study of patients' and health care providers' use of health-related internet information. J Med Internet Res 2015;17(6):e155 [FREE Full text] [doi: 10.2196/jmir.3792] [Medline: 26099267]
24. Yan M, Tan H, Jia L, Akram U. The antecedents of poor doctor-patient relationship in mobile consultation: a perspective from computer-mediated communication. Int J Environ Res Public Health 2020;17(7):2579 [FREE Full text] [doi: 10.3390/ijerph17072579] [Medline: 32283741]

## 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

*Edited by T de Azevedo Cardoso; submitted 14.03.23; peer-reviewed by Y Wang, E Assing Hvidt, C Beauvais; comments to author 31.05.23; revised version received 26.08.23; accepted 27.02.24; published 15.05.24.*

*Please cite as:*
*Nassehi D, Gripsrud BH, Ramvi E*
*Theoretical Perspectives Underpinning Research on the Physician-Patient Relationship in a Digital Health Practice: Scoping Review*
*Interact J Med Res 2024;13:e47280*
*URL: https://www.i-jmr.org/2024/1/e47280*
*doi:10.2196/47280*
*PMID:38748465*

Review

# Digital Health Literacy and Its Association With Sociodemographic Characteristics, Health Resource Use, and Health Outcomes: Rapid Review

Eva Yuen[1,2,3], PhD; Natalie Winter[1,2], PhD; Feby Savira[4,5,6], PhD; Catherine E Huggins[4,6], PhD; Lemai Nguyen[2,7], PhD; Paul Cooper[8], PhD; Anna Peeters[9], PhD; Kate Anderson[6,10], PhD; Rahul Bhoyroo[11], MPH, MD; Sarah Crowe[11], PhD; Anna Ugalde[1,2], PhD

[1]School of Nursing and Midwifery, Deakin University, Burwood, Australia

[2]Centre for Quality and Patient Safety Research, Institute for Health Transformation, Deakin University, Burwood, Australia

[3]Monash Health, Clayton, Australia

[4]Global Centre for Preventative Health and Nutrition, Institute for Health Transformation, Deakin University, Burwood, Australia

[5]Deakin Health Economics, Institute for Health Transformation, Deakin University, Burwood, Australia

[6]School of Health and Social Development, Deakin University, Burwood, Australia

[7]Department of Information Systems and Business Analytics, Deakin Business School, Deakin University, Burwood, Australia

[8]School of Medicine, Deakin University, Burwood, Australia

[9]Institute for Health Transformation, Deakin University, Burwood, Australia

[10]School of Computing Technologies, STEM College, RMIT University, Melbourne, Australia

[11]Western Victoria Primary Health Network, Geelong, Australia

**Corresponding Author:**
Eva Yuen, PhD
School of Nursing and Midwifery
Deakin University
221 Burwood Highway
Burwood, 3125
Australia
Phone: 61 03 92517074
Email: e.yuen@deakin.edu.au

## Abstract

**Background:** Digital health literacy has emerged as a critical skill set to navigate the digital age.

**Objective:** This review sought to broadly summarize the literature on associations between digital health literacy and (1) sociodemographic characteristics, (2) health resource use, and (3) health outcomes in the general population, patient groups, or parent or caregiver groups.

**Methods:** A rapid review of literature published between January 2016 and May 2022 was conducted through a search of 4 web-based databases. Articles were included on the basis of the following keywords: “measured digital health literacy,” “digital literacy,” “ehealth literacy,” “e-health literacy,” “electronic health literacy,” or “internet health literacy” in adult populations; participants were from countries where English was the primary language; studies had to be cross-sectional, longitudinal, prospective, or retrospective, and published in English.

**Results:** Thirty-six articles met the inclusion criteria. Evidence on the associations between digital health literacy and sociodemographic characteristics varied (27/36, 75% included studies), with higher education (16/21, 76.2% studies that examined the association) and younger age (12/21, 57.1% studies) tending to predict higher digital health literacy; however, other studies found no associations. No differences between genders were found across the majority of studies. Evidence across ethnic groups was too limited to draw conclusions; some studies showed that those from racial and ethnic minority groups had higher digital health literacy than White individuals, while other studies showed no associations. Higher digital health literacy was associated with digital health resource use in the majority of studies (20/36, 55.6%) that examined this relationship. In addition, higher digital health literacy was also associated with health outcomes across 3 areas (psychosocial outcomes; chronic disease and health management behaviors; and physical outcomes) across 17 included studies (17/36, 47.2%) that explored these relationships.

However, not all studies on the relationship among digital health literacy and health resource use and health outcomes were in the expected direction.

**Conclusions:** The review presents mixed results regarding the relationship between digital health literacy and sociodemographic characteristics, although studies broadly found that increased digital health literacy was positively associated with improved health outcomes and behaviors. Further investigations of digital health literacy on chronic disease outcomes are needed, particularly across diverse groups. Empowering individuals with the skills to critically access and appraise reliable health information on digital platforms and devices is critical, given emerging evidence that suggests that those with low digital health literacy seek health information from unreliable sources. Identifying cost-effective strategies to rapidly assess and enhance digital health literacy capacities across community settings thus warrants continued investigation.

***(Interact J Med Res 2024;13:e46888)*** doi:10.2196/46888

**KEYWORDS**

digital health literacy; eHealth literacy; health literacy; digital health; web-based database; health information

## Introduction

As health technologies evolve, digital devices, health-related apps, and web-based portals are increasingly used to deliver and access medical information and health care services [1]. While such technologies can be a gateway to health information and support [2], research also predicts a "digital divide" in which an individual's sociodemographic characteristics (eg, age, education, and income) influence their effective engagement with digital health information [3]. Digital health literacy has emerged as a vital skill set to navigate health care in the digital age [4]. Digital health literacy has been described as an extension of eHealth literacy, which captures the skills to seek, find, understand, and critically appraise health information from electronic sources to manage one's own health [5,6]. Digital health literacy has been posited to expand on the definition by emphasizing the individual as both an active participant and a distributer of digital health information, not just a passive receiver [6]. Digital health literacy has skills unique to health literacy, including computer literacy, media literacy, and critical appraisal skills to identify and evaluate reliable information and resources [1]. The importance of digital health literacy is increasingly recognized for its role in optimal individual and population health [4] and critical to limiting health inequalities [7].

Reviews have reported associations between digital health literacy and health outcomes across specific populations. Among people with long-term conditions, higher digital health literacy was predominantly associated with greater health-promoting behaviors [8]. A recent review identified that both older adults with cancer and their carers reported low digital health literacy and decreased confidence in appraising digital health information, with barriers identified including low socioeconomic status, poor digital access, and lack of familiarity and use [9]. Several reviews have examined the impact of digital health literacy on health outcomes in specific populations of interest (eg, older adults [9], those with long-term conditions [8], college students [10], and underserved [11] or vulnerable populations [12]). However, existing evidence on the relationships between digital health literacy, health outcomes, and sociodemographic characteristics across broad population groups has not been synthesized to date. Current literature on digital health literacy has focused on definitions and scales; its associations with health outcomes; the digital divide; and influencing factors of health literacy [13]. A review of associations between (1) sociodemographic characteristics, (2) health resource use, and (3) health outcomes in the general population, patient groups, or parent or caregiver groups is currently lacking. This information is critical to inform the development and implementation of digital health strategies to improve digital health literacy in communities with the highest need. This review sought to broadly summarize the literature on associations between digital health literacy and (1) sociodemographic characteristics, (2) health resource use, and (3) health outcomes.

## Methods

A rapid review was undertaken following the principles of a systematic review [14]; however, with some simplification of steps to ensure a timely and accurate synthesis of evidence to inform the development of a digital health strategy for implementation across community settings. Given the rapid nature of the review, the review and protocol were not registered with an international register.

A search for peer-reviewed publications was undertaken within the CINAHL, PsycINFO, MEDLINE, and Embase databases in May 2022. English language articles published between January 2016 and May 2022 were included. Given then rapid changes in technology and digital engagement observed in the health field, we sought to comprehensively review contemporary evidence from the preceding 5 years only. Additional searches were conducted in June 2022 using Google Scholar and by handsearching the reference lists of included papers to ensure that all relevant literature were captured in the review. Search terms synonyms were "digital health literacy," "e-health literacy," "electronic health literacy," "internet literacy," "internet health literacy," and "digital literacy." Key search terms are detailed in Multimedia Appendix 1. For the purposes of the review, studies that examined electronic health literacy or eHealth literacy were included given similarities in definitions of these concepts [13]. The inclusion and exclusion criteria are summarized in Table 1. In brief, studies that examined digital health literacy using a validated measure of digital health literacy in the general population, patient groups, or parent or caregiver groups were eligible for inclusion. Articles were excluded if

they focused exclusively on college students, since a review of this group has already been conducted [10].

**Table 1.** Inclusion and exclusion criteria.

| Inclusion | Exclusion |
|---|---|
| Assesses digital health literacy in the general population, patients, parents, or caregivers | Study population younger than 18 years |
| Sample included individuals from Australia, New Zealand, the United States, Canada, the United Kingdom, or Ireland | Primary language of country is not English |
| Quantitative study | Qualitative study, literature or systematic review, commentary, conference abstract, opinion piece, protocol, or thesis |
| Published in 2016 onwards | Non–health care–related (eg, high school education) |
| —[a] | Not focused on patients or general population (eg, digital health literacy of health providers) |
| — | Focusses on digital health literacy of interventions, programs, etc |
| — | Assesses psychometric properties of a digital health literacy measure |
| — | Measures digital health literacy using a single item or proxy measure |

[a]Not applicable.

The search results were imported into a review management dashboard, Covidence, to allow for simultaneous screening between reviewers. Four reviewers (EY, NW, FS, and FT) independently screened titles and abstracts, and the full texts were screened by 5 reviewers (EY, NW, FS, FT, and SR). At both screening stages, each paper was assessed by 2 reviewers, with discrepancies resolved through discussion. The screening process was reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15].

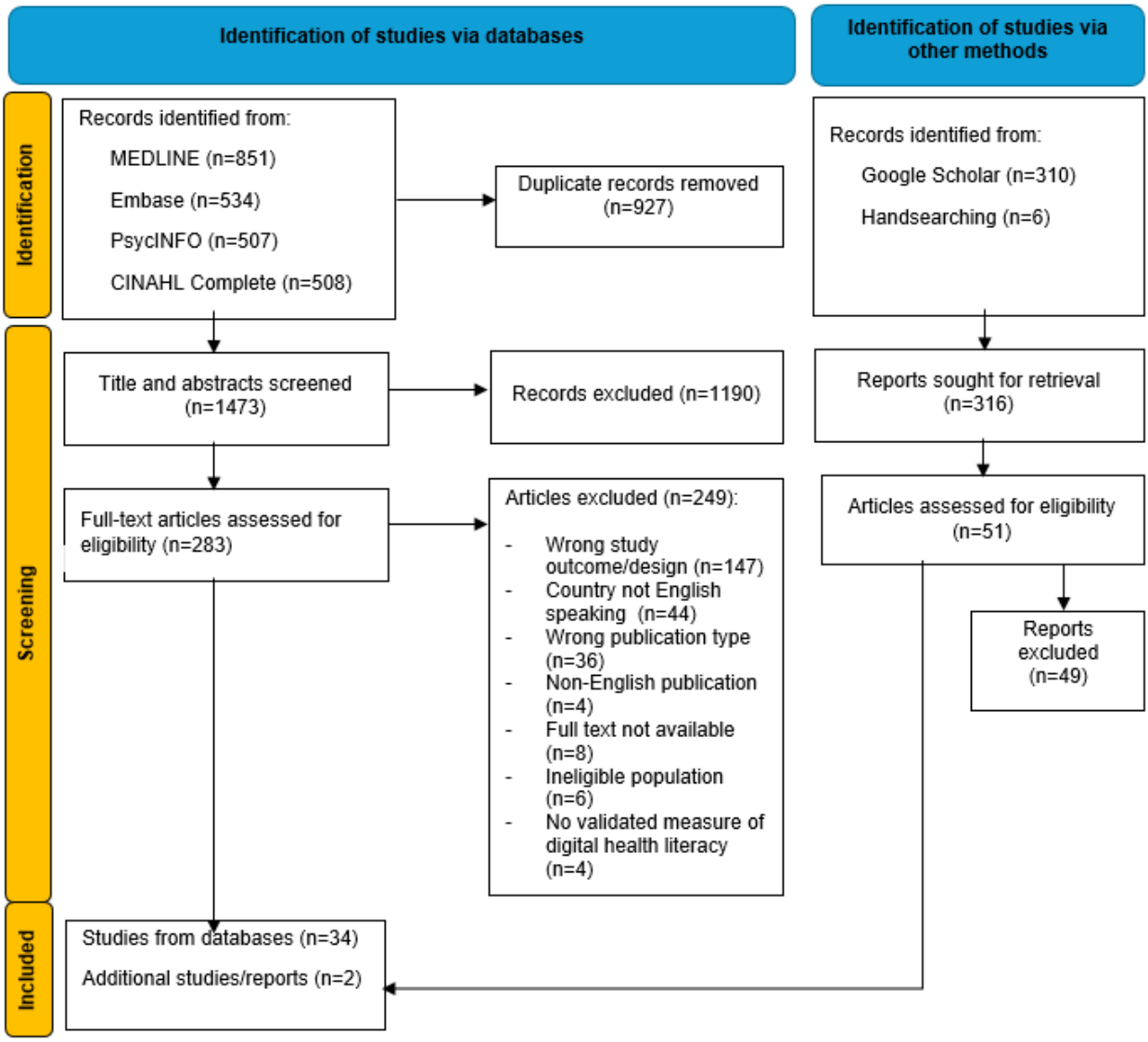
For studies included in the review, data were extracted on (1) study characteristics (eg, author, publication year, country, study design, sample size, participant characteristics, measures used to assess digital health literacy, and health literacy scores); (2) health outcomes associated with digital health literacy; and (3) sociodemographic characteristics associated with digital health literacy. Given the rapid nature of the review to inform strategy, data extraction for each article was conducted by 1 author only (EY, NW, or FS), with data extracted presented to the research team for review. Results were narratively synthesized to address the 3 aims and to group findings into identified themes.

## Results

### Description of Included Studies

Of the 1473 articles, 36 met the inclusion criteria (Figure 1). Studies are summarized in Multimedia Appendix 2. Twenty-five studies were from the United States, 4 studies from Australia [16-19], 4 from Canada [20-23], 1 from the United Kingdom [24], and 1 cross-cultural study that included people from Australia and India [25]. Studies included 34 cross-sectional survey designs, 1 mixed methods study [26], and 1 randomized controlled trial [27]. Sample sizes ranged from 22 [28] to 3258 [29] participants. The majority of studies (32/36, 88.9%) assessed digital health literacy using the eHealth Literacy Scale [30]. Across studies, variations on how digital health literacy scores were reported; some used a mean score (with higher scores indicating higher digital health literacy) and other studies used cutoff scores to determine those in high compared with low digital health literacy categories.

**Figure 1.** PRISMA Flowchart.




Five distinct population groups were identified across the studies. These included the general population, including older adults, patient groups, minority populations, and caregivers. Patient groups included people with HIV, cancer, chronic obstructive pulmonary disease (COPD), spinal injury, kidney disease, bipolar disorder, otolaryngological disease, orthopedic trauma, cardiovascular risk, diabetes, and transplant recipients. Minority groups included African Americans, Southeast Asians, adults in rural or remote regions, low-income groups, transgender and gender diverse, and young men who have sex with men. Caregivers included those of children with special health care needs, men with prostate cancer, and pediatric inpatients with subacute health conditions.

## Associations Between Digital Health Literacy and Sociodemographic Characteristics

Twenty-seven (75%) studies were identified as examining associations between digital health literacy and sociodemographic characteristics, with varied findings. These are shown in Table 2 and Figure 2 and discussed in the following sections.

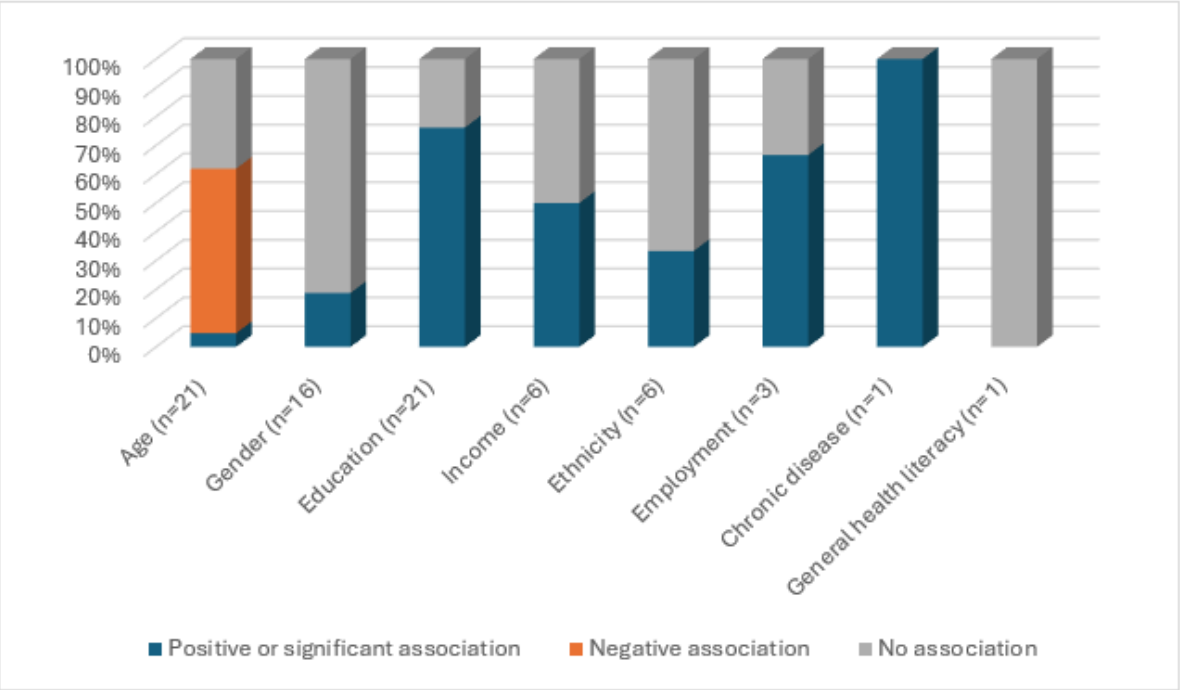
**Table 2.** Digital health literacy and associations with sociodemographic characteristics (N=27).

| Sociodemographic characteristics and association with digital health literacy | Articles in which the finding occurred, n | | | | | |
|---|---|---|---|---|---|---|
| | Patient group | General population | Minority group | Older adults | Caregivers | Total |
| **Age (years)** | | | | | | |
| Negative | 5 | 3 | 1 | 1 | 2 | 12 |
| Positive | 1 | —[a] | — | — | — | 1 |
| None | 5 | — | 1 | 2 | — | 8 |
| **Gender**[b] | | | | | | |
| Significant | 1 | 1 | 1 | — | — | 3 |
| None | 7 | 3 | — | 3 | — | 13 |
| **Education** | | | | | | |
| Positive | 7 | 3 | 1 | 3 | 2 | 16 |
| None | 3 | 1 | — | 1 | — | 5 |
| **Income** | | | | | | |
| Positive | 1 | 1 | — | — | 1 | 3 |
| None | 2 | — | — | 1 | — | 3 |
| **Race/ethnicity** | | | | | | |
| Positive | — | 2 | — | — | — | 2 |
| None | 3 | — | — | — | 1 | 4 |
| **Employment** | | | | | | |
| Positive | 1 | 1 | — | — | — | 2 |
| None | — | — | — | 1 | — | 1 |
| **No chronic disease** | | | | | | |
| Positive | — | — | 1 | — | — | 1 |
| **General health literacy** | | | | | | |
| None | — | — | — | 1 | — | 1 |
| **Marital status** | | | | | | |
| None | 3 | 1 | — | 2 | — | 6 |
| **Languages spoken** | | | | | | |
| None | 1 | 1 | — | — | — | 2 |
| **Socioeconomic status** | | | | | | |
| None | 1 | — | — | 1 | — | 2 |
| **Rural/urban** | | | | | | |
| None | 1 | — | — | — | — | 1 |

[a]Not applicable.

[b]Women had higher digital health literacy.

**Figure 2.** Digital health literacy and associations with sociodemographic characteristics. Positive association: higher digital health literacy associated with higher outcome of characteristics (eg, higher digital health literacy associated with higher income); negative association: higher digital health literacy associated with lower characteristics (eg, higher digital health literacy associated with younger age).




### *Age*

Twenty-one studies (21/36, 58.3%) explored associations between age and levels of digital health literacy with mixed results. Twelve studies (12/21, 57.1%) reported a negative association between age and digital health literacy, with older people more likely to have lower digital health literacy than their younger counterparts [16,19-21,24-26,31-35]. Negative associations between age and digital health literacy were found in the general population (3/21, 14.3%) [16,25,32], as well as the following specific population studies: Southeast Asian adults in Canada [21]; older adults [19,20]; people with chronic kidney disease [26], breast cancer [24], or cardiovascular risk [19]; and caregivers [34,35]. One study that used the eHealth Literacy Questionnaire [16] found that age was negatively associated with 4 of 7 digital health literacy subscales (ability to engage with digital services; using technology to process health information; motivated to engage with digital health services; and digital services that suit individual needs).

In contrast, studies (1/21, 4.8%) found a positive significant association between age and digital health literacy in people with bipolar disorder, where older age significantly predicted higher digital health literacy scores [36]. Notably, the mean age in this sample was less than that in other studies, and more than three-quarters had completed some form of higher education. Furthermore, studies (8/21, 38.1%) reported no associations between age and digital health literacy across patient [17,23,37-39], rural [22], and older [40,41] populations.

### *Gender*

Digital health literacy and gender were reported in studies (16/36, 44.4%), with most (13/16, 81.3%) finding no gender differences [16,19,20,23,26,31-33,36,39-42]. The 3 studies that found that women were more likely to perceive themselves as having higher digital health literacy than men were from the general population in the United States [43], adults from rural communities in British Columbia with a population of 12,000 [22], and among otolaryngology patients [37].

### *Education*

Twenty-one (21/36, 58.3%) studies examined associations between digital health literacy and education levels, with two-thirds (16/21, 76.2%) finding higher digital health literacy positively associated with higher education levels [16,18-21,24,31-36,39,41,44,45]. Five studies (5/16, 31.3%) reported no associations between digital health literacy and education, including 4 studies of patient groups [21,28,29], 1 general population study [25], and 1 study of older adults [32]. Across these studies, large proportions of participants had greater than high school education levels (35%-84.8%).

### *Income*

Six studies (6/36, 28.6%) examined associations between digital health literacy and income, with half reporting no associations in people living with chronic illness [19,23] or those in a general population [41]. By contrast, the 3 studies that found significant positive associations between higher digital health literacy and income reported these findings in people with chronic illness [26], among caregivers [35], and in minority populations [32].

### *Ethnicity*

Six studies (6/36, 16.7%) examined associations between digital health literacy and ethnicity, with 2 reporting differences between ethnicity groups [25,32]. Black or African Americans had higher digital health literacy than their Caucasian counterparts in a stratified US sample [32]. Australians had higher digital health literacy than Indians in a cross-cultural

study [25]. Conversely, 4 studies found no differences in digital health literacy levels across ethnicity groups among those with chronic illness [26], spinal cord injury [38], cancer survivors [24], or caregivers of people with prostate cancer [35].

### *Employment*

Of the 3 studies (3/36, 8.3%) that examined associations between digital health literacy and employment, 1 patient group study (spinal cord injury [23]) and 1 general population study [43] found higher digital health literacy among those who were employed versus those who were unemployed. By contrast, another study found no differences in employment groups and digital health literacy in an older US sample [40].

### *Chronic Disease or Comorbidities*

One study (1/36, 2.7%) reported small but significant findings for higher digital health literacy in people with no chronic conditions, compared with those with a chronic condition, in a sample of South Asians in Canada [21].

### *General Health Literacy*

Only 1 study (1/36, 2.7%) examined associations between digital health literacy and general health literacy, finding no associations in an older adult sample [40].

### *Marital Status, Language Spoken, Socioeconomic Status, Household Size, Rural or Urban Location, or Country of Birth*

No associations between digital health literacy and the following sociodemographic characteristics were found: marital status [17,25,38,41,46,47], language spoken [16,20], socioeconomic status or deprivation [24,40], household size [40], or rural or urban location [24].

## Associations Between Digital Health Literacy and Health Resource Navigation

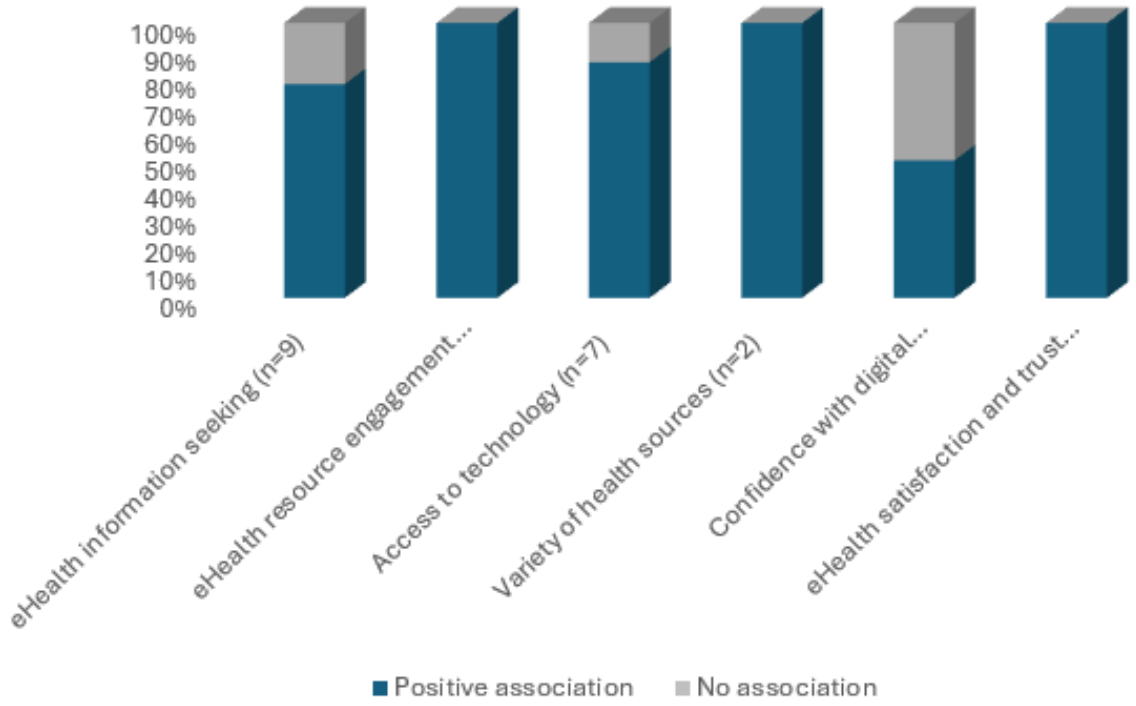
Twenty studies (20/36, 55.6%) examined associations between digital health literacy and health resource navigation distributed across all participant categories. Categories of health information engagement included eHealth information seeking; eHealth behaviors; access to and use of e-resources; use of health information sources; and eHealth satisfaction, and are shown in Table 3, Figure 3, and in the following sections.

**Table 3.** Digital health literacy and digital health resource navigation (N=17).

| Health outcomes | Association with digital health literacy | Articles in which association occurred, n | | | | | |
|---|---|---|---|---|---|---|---|
| | | Patient group | General population | Minority group | Older adults | Caregivers | Total |
| eHealth information seeking | Positive | 1 | 2 | 2 | 2 | —[a] | 7 |
| eHealth information seeking | None | — | — | 2 | — | — | 2 |
| eHeath resource engagement | Positive | 2 | 1 | 2 | 1 | 1 | 7 |
| Access to and use of technology | Positive | 1 | 1 | 2 | 1 | 2 | 7 |
| Use of a variety of health information sources | Positive | — | — | 1 | 1 | — | 2 |
| Confidence or comfort with using digital resources | Positive | — | — | — | 1 | — | 1 |
| Confidence or comfort with using digital resources | None | — | 1 | — | — | — | 1 |
| eHealth satisfaction and trust | Positive | — | 1 | 1 | — | 1 | 3 |

[a]Not applicable.

**Figure 3.** Digital health literacy and associations with health resource navigation.




### *Digital Health Information Seeking*

Nine studies (9/36, 25%) examined digital health literacy and digital health information seeking, and of these, 7 showed that higher digital health literacy was significantly positively associated with e-information seeking. These included studies of general populations [47,48], middle-aged to older adults [41,49], minority populations [18,21], and of people with HIV [50]. In addition, higher digital health literacy was associated with greater exposure to medical and health websites in a US general population [48] and among Southeast Asians in Canada [21]. Digital health literacy was also associated with eHealth information consumerism in older adults [41].

By contrast, 2 studies (2/36, 5.6%) found no associations between digital health literacy and eHealth information–seeking behaviors. In a sample of older Hispanic people in the United States, digital health literacy was not associated with use of, nor willingness to use the internet for health information [51]. A study of transgender and gender diverse people found no interactions between digital health literacy and web-based health-seeking behaviors [29].

### *eHealth Resource Engagement*

Ten studies (10/36, 27.8%) examined associations between digital health literacy and eHealth resource engagement with 7 studies reporting positive associations [18,21,31,33,34,36,47]. Higher digital health literacy was associated with signing up for email updates, watching health-related videos, seeking resources from people with similar lived experience, and using health indicator tracking in a general population [47] and among caregivers [34]. Digital health literacy was also associated with willingness to participate in mobile health research interventions, wearing a smart watch or tracking device, and downloading a health app among minority race or ethnicity groups [18,21]. Higher digital health literacy was associated with increased eHealth resource use in older adults [33] and people living with chronic illness [36] and greater use of patient portals in people with chronic disease [47], older adults [33], and transplant recipients [31]. However, no differences were found among varying digital health literacy levels and willingness to participate in research that used web-based forums, support groups, or counseling in an African American sample [18].

### *Access to and Use of Technology*

Seven studies (7/36, 19.4%) examined digital health literacy and associations with access to and use of technology with mostly positive associations reported. Access to the internet for personal use was associated with higher digital health literacy among caregivers [35]. Greater internet use was also associated with higher digital health literacy in a low-income population in the United States [52], in Southeast Asians in Canada [21], and in caregivers of children with special needs [34]. Access to and use of digital devices and the internet were associated with higher digital health literacy in an Australian population [16]. Access to any mobile device was also associated with higher eHealth literacy among breast cancer survivors [24]. Furthermore, among older adults, those who owned 2 or more electronic devices had higher digital health literacy than those with 1 or no devices, although no differences in internet use were found across digital health literacy levels [40].

### *Use of a Variety of Health Information Sources*

Two studies (2/36, 5.6%) examined digital health literacy and the use of a variety of health information sources. High digital health literacy was associated with the use of more health information sources compared with those with low digital health literacy among older adults [40] and among Black or African Americans [18]. Health information sources included the

internet, health books and magazines, TV programs, literature in medical offices, and discussions with health care providers [40]. This study also probed the relationship between digital health literacy and the sources of information used. People with high digital health literacy were more likely to rely on doctors' knowledge for medical decision making and drew upon more sources of health information than those with low digital health literacy [40]. Similarly, among Black or African Americans, higher digital health literacy was also significantly associated with citing the internet, nurses, books, radio, or news apps as sources of health information [18].

### *Confidence or Comfort With Using Digital Resources*

Two studies (2/36, 5.6%) examined confidence in using digital resources and digital health literacy. Older adults with higher digital health literacy were more likely to experience no stress when using a computer than those with lower digital health literacy [40]. By contrast, among population groups digital health literacy was not associated with comfort in using the internet [25].

### *eHealth Satisfaction and Trust*

Three studies (3/36, 8.3%) examined digital health literacy and associations with eHealth satisfaction and trust with positive outcomes reported. Higher digital health literacy was associated with greater telemedicine satisfaction among rural and remote communities [22]. Increased digital health literacy was also associated with greater positive perceptions of eHealth in caregivers [34].

In addition, higher digital health literacy was associated with greater perceived trust in eHealth information in a sample of Black or African Americans and Caucasians after controlling for socioeconomic status and social media use [32]. Notably, among individuals with low digital health literacy, older adults had higher perceived trust in Facebook and less trust in support groups than their younger counterparts [32]. Furthermore, in those with low digital health literacy, Black or African Americans were more likely to report greater perceived trust in web-based blogs or diaries and twitter than their younger or Caucasian counterparts; by contrast, for those with high digital health literacy, Black or African Americans were more likely than Caucasians to trust support groups [32].

## Associations Between Digital Health Literacy and Health Outcomes

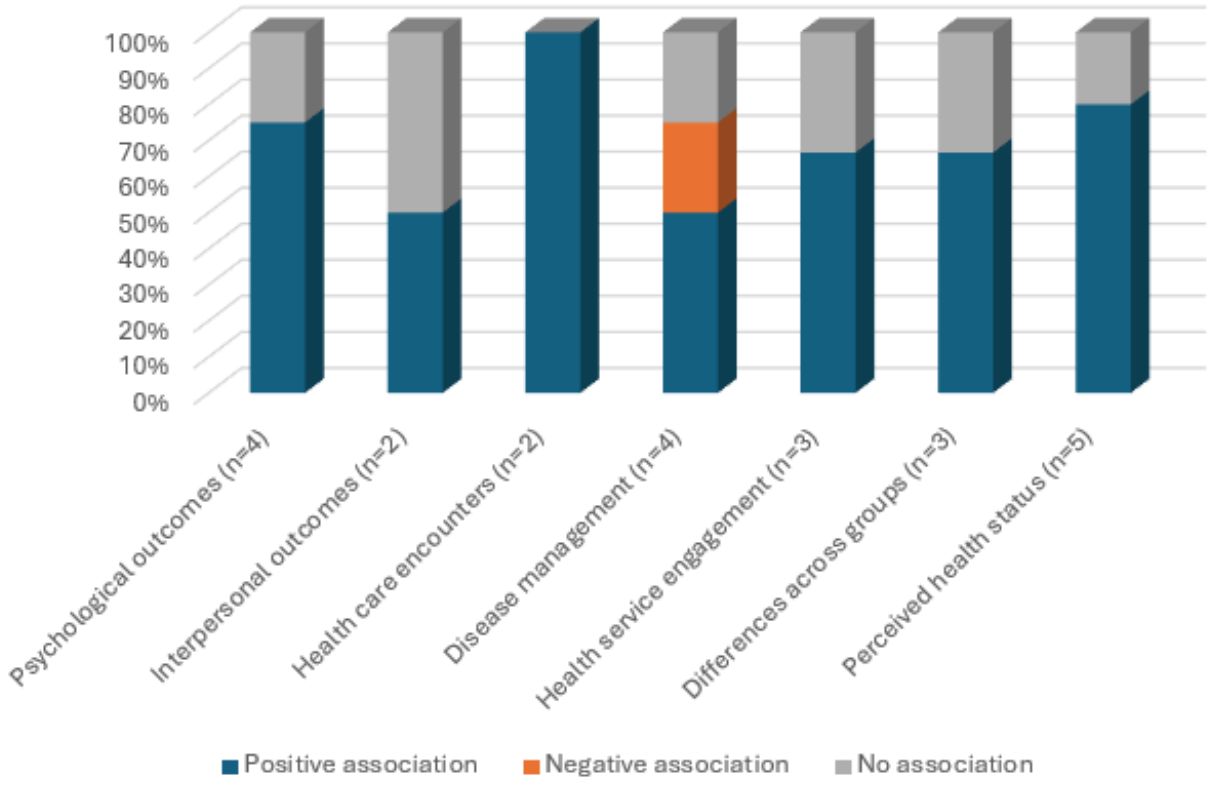
Seventeen studies (17/36, 47.2%) examined associations between digital health literacy and health outcomes. The outcomes are grouped into 3 main categories: psychosocial health outcomes, chronic disease and health management behavioral outcomes, and perceived health status. These are shown in Table 4 and Figure 4 and described in the following sections.

**Table 4.** Digital health literacy and health outcomes.

| Category and subcategory of outcomes, and association with digital health literacy | Articles in which the association occurred, n | | | | | |
|---|---|---|---|---|---|---|
| | Patient group | General population | Minority group | Older adults | Caregivers | Total |
| **Psychological** | | | | | | |
| **Psychological** | | | | | | |
| Positive | —[a] | 1 | — | 2 | — | 3 |
| None | — | — | — | — | 1 | |
| **Interpersonal** | | | | | | |
| Positive | — | — | — | — | 1 | 1 |
| None | — | — | — | — | 1 | 1 |
| **Satisfaction with health care encounters** | | | | | | |
| Positive | — | — | — | 2 | — | 2 |
| **Behaviors for managing health or chronic disease** | | | | | | |
| **Disease self-efficacy and self-management** | | | | | | |
| Positive | 1 | — | 1 | — | 1 | 3 |
| None | — | — | 1 | — | — | 1 |
| **Health risk behaviors** | | | | | | |
| Positive | — | — | 1 | — | — | 1 |
| **Health service engagement** | | | | | | |
| Positive | — | — | 1 | 1 | — | 2 |
| None | 1 | — | — | — | — | 1 |
| **Health status across patient groups** | | | | | | |
| Positive | 3 | — | — | — | — | 3 |
| **Perceived health status** | | | | | | |
| Positive | 2 | — | 1 | 1 | — | 4 |
| None | — | 1 | — | — | — | 1 |

[a]Not applicable.

**Figure 4.** Digital health literacy and behaviors for managing health and chronic disease.




### *Psychosocial Health Outcomes*

Digital health literacy and its relationship with psychosocial health outcomes included diverse subcategories related to psychological outcomes, interpersonal factors, and satisfaction with health care encounters.

#### Psychological Outcomes

Four studies (4/36, 11.1%) examined associations between digital health literacy and psychological outcomes. Higher digital health literacy was significantly associated with greater empowerment through information seeking [49] and less affective distress [44] in older adult populations. Higher digital health literacy was also related to increased information seeking, which was associated in lower cancer fatalism (ie, inevitable death following a cancer diagnosis) in a general population from the United States [48]. By contrast, no associations were found between digital health literacy and worry among caregivers of children with special needs [34].

#### Interpersonal Outcomes

Two studies (2/36, 5.6%) examined relationships between digital health literacy and interpersonal outcomes. One study found that digital health literacy was significantly associated with the size of social networks for seeking information and support for health decision-making in a sample of caregivers of people with prostate cancer [35]. However, digital health literacy was not associated with social functioning, family relationships, or communication skills among caregivers of children with special needs [34].

#### Satisfaction With Health Care Encounters

Two studies (2/36, 5.6%) examined associations between digital health literacy and health care encounters among older adults. Higher digital health literacy was significantly associated with greater satisfaction [41] and less perceived strain in medical encounters [44] in older populations.

### *Behaviors for Managing Health or Chronic Disease*

Digital health literacy and individuals' management of health or chronic disease were related to (1) disease self-efficacy, disease management, and health risk behaviors; (2) health service engagement; and (3) health status across patient groups.

#### Disease Management and Self-Efficacy

Four studies (4/36, 11.1%) examined digital health literacy and disease management and self-efficacy. Three studies examined associations between digital health literacy and individual management of chronic disease and health conditions, with mixed results. In caregivers of men with prostate cancer, higher digital health literacy was associated with greater likelihood of getting a second opinion, awareness of treatment options, and size of social network for information and support in treatment decision-making [23]. In a study of young men who have sex with men, low digital health literacy was associated with decreased likelihood of evaluating personal risk for HIV/STIs, educating others about HIV/STIs, and getting tested for HIV/STI after completing an HIV/STI education intervention [27]. By contrast, in a sample of transgender and gender diverse people [29], no relationship between self-reported digital health literacy and adherence to human papillomavirus vaccination was found.

In 1 study of people living with COPD, those with higher digital health literacy reported greater self-efficacy with managing their chronic disease [39]. In another study that examined associations between digital health literacy and associations with health risk behaviors, higher digital health literacy was associated with greater HIV transmission risk behaviors (eg, unprotected sexual activities or illicit drug use) among women infected with HIV [50].

#### Health Service Engagement

Three studies (3/36, 17.6%) examined associations between digital health literacy and health service engagement with positive findings reported in 2 studies. Higher digital health literacy was associated with greater number of general practitioner visits through increased searches for health information among older adults [49]. Higher digital health literacy levels were also associated with greater likelihood of attending a physical examination by a physician in the prior 12 months in a Black or African Americans sample [32]. By contrast, 1 study found no associations between digital health literacy and health service engagement in a study of transplant recipients. Participants with low digital health literacy were less likely to have talked with a doctor about injury information than those with higher digital health literacy [31].

#### Differences in Digital Health Literacy Across Patient Groups and Health Status

Three studies (3/36, 17.6%) compared digital health literacy across different patient or health status groups and identified varying levels of digital health literacy. Among low-income pregnant women with gestational diabetes or type II diabetes, those with gestational diabetes trended toward higher digital health literacy than those with type II diabetes [28]. In another study, kidney transplant recipients were found to have higher digital health literacy than liver transplant patients [31]. Another paper [46] found higher digital health literacy levels among people with very severe COPD than among those with less severe COPD. In addition, those with lower lung–specific health-related quality of life also had higher digital health literacy levels [46]. These findings were attributed to those with more severe disease accessing the e-resources more frequently to find strategies and information on how to self-manage their disease [46].

#### *Perceived Health Status*

Five studies (5/36, 13.9%) examined associations between digital health literacy and perceived health status, with studies broadly reporting positive associations. Higher digital health literacy was significantly associated with higher self-reported health status among veterans with spinal cord injury [38] and among African Americans [18]. Digital health literacy was also associated with better self-care, perceived improved quality of life, and increased health status in a community sample aged between 40 and 93 years [41]. Similarly, higher digital health literacy was associated with higher scores on mobility, self-care and usual activities, and lung-specific health-related quality of life in people with COPD; however, digital health literacy was not associated with generic health-related quality of life in this sample [46]. By contrast, no associations were found between digital health literacy and health status in a sample of older US adults [40].

#### *Physical or Neurocognitive Health Outcomes*

One study each (2/36, 5.6%) explored digital health literacy on physical [38] or neurocognitive health outcomes [45]. Among veterans with spinal cord injury or disorder, no associations were found between digital health literacy and level or duration of injury [38]. In people with HIV, lower neurocognitive function was moderately associated with lower digital health literacy scores [45].

## Discussion

### Principal Findings

This review advances our understanding of consumer digital health literacy through identifying and synthesizing recent literature that explored digital health literacy and its relationship with sociodemographic characteristics and health outcomes. The findings present mixed results regarding the relationship between digital health literacy and sociodemographic characteristics. However, studies broadly suggested that increased digital health literacy was positively associated with improved health outcomes. Based on the findings, we derive implications for practice in closing the digital health divide.

Education was the most common characteristic associated with digital health literacy. People with higher education levels were more likely to have increased literacy skills to better read and interpret web-based health information. Moreover, education has been shown to be a predictor of use of eHealth resources [16]. Education is considered a major determinant of *health literacy*, since educational levels often influence literacy skills, employment status and income, and, as such, enable access to better living circumstances and access to health care [53]. However, some individuals with higher education levels may still have inadequate health literacy [54]. Examining specific skills and knowledge, rather than sociodemographic characteristics alone, may offer a more comprehensive understanding of this relationship in future [53].

Some studies showed that older individuals were more likely to have lower digital health literacy than their younger counterparts, although other studies found no associations. Research has shown that digital health literacy decreases with age, and this is explained by age-related cognitive changes, decreased vision and hearing, reduced motor functioning, and decreased health status [55]. Notably, in our current review, studies with patient populations found no associations between age and digital health literacy levels, and 1 recent study [36] with older people had higher digital health literacy scores than their younger counterparts. Furthermore, in this study [36] the sample was younger and highly educated, and the findings could be related to increased use of digital products. People with chronic health conditions are likely required to engage with digital health resources to manage their health condition, which may account for their higher digital health literacy levels. Thus, while some findings suggest that older people require additional supports to improve digital health literacy capacities, given the varied outcomes, more in-depth analysis of these relationships

between digital health literacy and age is recommended for future research including a full systematic review and meta-analysis. Findings showed predominantly no differences in digital health literacy across gender. While several studies reported that women had higher digital health literacy levels than men, the majority found no associations, suggesting that gender is not a likely predictor of digital health literacy. These findings may be attributed to increased use of electronic devices across populations and the closing gap in education levels between men and women [16].

Of the few included studies that examined ethnicity and digital health literacy, findings were also mixed. While some studies reported higher digital health literacy in White populations than in minority groups [25,35], other studies found no differences across ethnic groups [26,38], or conversely that people from minority groups had higher digital health literacy levels [24,32]. Findings from the current review suggest that ethnicity is not a reliable predictor of digital heath literacy; however, further evidence is needed. In addition, studies on digital health literacy among culturally and linguistically diverse communities were limited to a qualitive study, which was excluded from the current review [56]. Quantitative studies in culturally and linguistically diverse populations that met the study criteria were lacking. Thus, further studies are needed to explore these associations and to identify facilitators and barriers to digital health literacy for people in culturally and linguistically diverse communities.

The majority of studies reported positive associations between digital health literacy and health resource navigation, including higher levels of e-information seeking, e-resource engagement, access and use of technology, engagement with health information sources, eHealth satisfaction and trust, and comfort and use of digital resources. The findings suggest a bidirectional relationship between digital health literacy and eHealth resource use: those with greater access to digital devices and greater use of the internet had higher digital health literacy. It is likely that those with more confidence engaging with digital products will be more inclined to seek out health information in digital formats. The findings suggest that ensuring access to technology, as well as fostering skills to engage with eHealth resources is essential to promoting digital health literacy. Furthermore, for people with limited access to the internet or devices, we recommend providing information in nondigital formats.

Our review found that studies examined associations between digital health literacy and health outcomes including psychosocial outcomes, behaviors for managing health or chronic disease, and perceived health status. Studies showed that higher digital health literacy was associated with greater satisfaction with medical encounters, less perceived strain in medical encounters, increased empowerment, greater social networks for health information, and reduced affective distress and cancer fatalism. In these studies, higher digital health literacy enabled individuals to seek and understand information, which enabled them to feel more empowered to manage their health. However, findings were less consistent for associations between digital health literacy and chronic disease and health management outcomes.

Studies of chronic disease populations showed that digital health literacy was associated with greater disease knowledge and increased disease management efficacy, consistent with a prior systematic review [57]. However, while some studies suggested that digital health literacy was associated with increased health decision-making behaviors and health service use, other studies found no associations with health promotion behaviors or health service use.

Our findings support conclusions drawn from existing reviews [8], which identified that more studies are needed in the digital health field to examine whether digital advances are facilitating better outcomes for those with greater skills in using e-resources for health purposes. Furthermore, our findings from the review also highlight gaps in recent evidence on the impacts of digital health literacy on prevalent chronic health conditions (eg, cancer, heart disease, diabetes, stroke, and COPD). Given the increased use of telehealth during the COVID-19 pandemic, and importance of chronic disease self-management, research to understand how digital health literacy influences a person's capacity to engage with digital health resources to manage their health is needed. Furthermore, research on understanding the impacts of digital health literacy on health outcomes in chronic disease populations is recommended. In addition, only 2 included studies examined digital health literacy in caregiver populations. Caregivers, particularly those of adult care recipients, may have unique health information needs given their role in providing support. Given the evidence gap, we recommend further research on caregivers to identify their digital health literacy needs and how these can be addressed across health settings.

## Study Limitations

Across all studies, digital health literacy was assessed using self-reported measures (eg, eHealth Literacy Scale) rather than assessments of specific digital knowledge and skills. Thus, these perceived digital health literacy skills may not translate to everyday digital health behaviors. Studies also varied in their reports of digital health literacy scores (eg, means vs cutoff criteria), thus limiting comparability across studies. Furthermore, included studies captured digital health literacy levels across a range of participant groups, which limit in-depth understanding of digital health literacy within specific populations. Thus, findings should be interpreted with caution for specific groups. Only bivariate associations were reported in the current review. Thus, nuanced understanding of relationships between digital health literacy and health outcomes and sociodemographic characteristics may be excluded. In addition, the aim of the rapid review was to inform the development and implementation of emerging digital health strategies across community settings in real time. In contrast to a full systematic review, the rapid review approach trades some methodological rigor for efficiency in addressing a critical topic and may therefore be vulnerable to bias. While MeSH terms were not used in the database search to limit outputs to articles that focused on measuring digital health literacy, ~1500 publications were still identified. Finally, the findings were limited to studies conducted from English-speaking countries. Thus, articles that capture key findings of the review may have been excluded. Notwithstanding, we note that the current findings highlight

substantial gaps in research pertaining to digital health literacy within English-speaking countries and identify areas for future investigation.

## Conclusions

Findings from this review suggest that sociodemographic characteristics may predict digital health literacy levels in some but not all contexts, but evidence suggests that these are not deterministic. Although in its infancy and with limited evidence, studies show some associations between increased digital health literacy and various improved health behaviors and outcomes. Further investigations of digital health literacy on positive chronic disease outcomes are needed, particularly across underrepresented but key populations, including diverse cultural and chronic disease groups. Empowering individuals with the skills to critically access and appraise reliable health information on digital platforms and devices is vital, given emerging evidence that suggests that those with low digital health literacy seek health information from unreliable sources. Identifying cost-effective strategies to rapidly assess and enhance digital health literacy capacities across community settings thus warrants continued further investigation. Our findings also confirm a warning that those with greatest digital skills may obtain greatest benefit from access to digital health resources and vice versa, with the implication that digital divides (gaps between knowledge of digital skills and access to health information) may become entrenched without specific efforts to overcome such divides. Our review contributes to the global digital health movement by identifying areas that require further investigation. It emphasizes the pivotal role of digital health literature in improving health care outcomes and promoting a more inclusive health care system. Digitalization and digital technologies transform and enhance the delivery of health care services; therefore, digital health literature becomes essential to engage health consumers and empower them to actively participate in their own health care and address health inequalities.

## Acknowledgments

The authors received funding from Western Victoria Primary Health Network. EY is supported by a Deakin University Dean's Health Research Fellowship. FS is supported by a Deakin University Alfred Deakin Postdoctoral Research Fellowship. AU is supported by a Victorian Cancer Agency Early Career Health Services Fellowship. AP is supported by a National Health and Medical Research Council Investigator Grant. The authors would like to sincerely thank the following members of the research team for their contributions to the article screening: Fisaha Tesfay, Sharina Riva, and Joanne Adams.

## Data Availability

Data collection forms and extracted data are available upon reasonable request.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search strategies.
[DOCX File , 19 KB - ijmr_v13i1e46888_app1.docx ]

Multimedia Appendix 2
Characteristics of included studies (n=34).
[DOCX File , 38 KB - ijmr_v13i1e46888_app2.docx ]

## References

1. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
2. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
3. Watts G. Covid-19 and the digital divide in the UK. Lancet Digit Health 2020;2(8):e395-e396 [FREE Full text] [doi: 10.1016/S2589-7500(20)30169-2] [Medline: 32835198]
4. Kim J, Jin B, Livingston MA, Henan MD, Hwang J. Fundamentals of digital health literacy: a review of the literature. 2022 Presented at: Adult Education Research Conference; 2022; Norman, OK.
5. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
6. Dong Q, Liu T, Liu R, Yang H, Liu C. Effectiveness of digital health literacy interventions in older adults: single-arm meta-analysis. J Med Internet Res 2023;25:e48166 [FREE Full text] [doi: 10.2196/48166] [Medline: 37379077]

7. van Kessel R, Wong BLH, Clemens T, Brand H. Digital health literacy as a super determinant of health: more than simply the sum of its parts. Internet Interv 2022;27:100500 [FREE Full text] [doi: 10.1016/j.invent.2022.100500] [Medline: 35242586]
8. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patients with long-term conditions. Eur Psychol 2019;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
9. Verma R, Saldanha C, Ellis U, Sattar S, Haase KR. eHealth literacy among older adults living with cancer and their caregivers: a scoping review. J Geriatr Oncol 2022;13(5):555-562. [doi: 10.1016/j.jgo.2021.11.008] [Medline: 34810146]
10. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth literacy among college students: a systematic review with implications for eHealth education. J Med Internet Res 2011;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
11. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
12. Choukou MA, Sanchez-Ramirez DC, Pol M, Uddin M, Monnin C, Syed-Abdul S. Covid-19 infodemic and digital health literacy in vulnerable populations: a scoping review. Digit Health 2022;8:20552076221076927 [FREE Full text] [doi: 10.1177/20552076221076927] [Medline: 35223076]
13. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022;24(7):e35816 [FREE Full text] [doi: 10.2196/35816] [Medline: 35793141]
14. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012;1(1):1-9 [FREE Full text] [doi: 10.1186/2046-4053-1-10] [Medline: 22587960]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264-269 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
16. Cheng C, Elsworth G, Osborne RH. Validity evidence based on relations to other variables of the eHealth literacy Questionnaire (eHLQ): bayesian approach to test for known-groups validity. J Med Internet Res 2021;23(10):e30243 [FREE Full text] [doi: 10.2196/30243] [Medline: 34647897]
17. Hyde LL, Boyes AW, Mackenzie LJ, Leigh L, Oldmeadow C, Riveros C, et al. Electronic health literacy among magnetic resonance imaging and computed tomography medical imaging outpatients: cluster analysis. J Med Internet Res 2019;21(8):e13423 [FREE Full text] [doi: 10.2196/13423] [Medline: 31464188]
18. James DCS, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mHealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
19. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
20. Cherid C, Baghdadli A, Wall M, Mayo NE, Berry G, Harvey EJ, et al. Current level of technology use, health and eHealth literacy in older Canadians with a recent fracture-a survey in orthopedic clinics. Osteoporos Int 2020;31(7):1333-1340. [doi: 10.1007/s00198-020-05359-3] [Medline: 32112118]
21. Makowsky MJ, Davachi S, Jones CA. eHealth literacy in a sample of South Asian adults in Edmonton, Elberta, Canada: subanalysis of a 2014 community-based survey. JMIR Form Res 2022;6(3):e29955. [doi: 10.2196/29955] [Medline: 35353044]
22. Rush KL, Seaton C, Li E, Oelke ND, Pesut B. Rural use of health service and telemedicine during Covid-19: the role of access and eHealth literacy. Health Informatics J 2021;27(2):14604582211020064. [doi: 10.1177/14604582211020064] [Medline: 34041936]
23. Singh G, Sawatzky B, Nimmon L, Ben Mortenson W. Perceived eHealth literacy and health literacy among people with spinal cord injury. Arch Phys Med Rehabil 2021;102(10):e50. [doi: 10.1016/j.apmr.2021.07.614]
24. Moon Z, Zuchowski M, Moss-Morris R, Hunter MS, Norton S, Hughes LD. Disparities in access to mobile devices and e-health literacy among breast cancer survivors. Support Care Cancer 2022;30(1):117-126 [FREE Full text] [doi: 10.1007/s00520-021-06407-2] [Medline: 34236506]
25. Austin DW, Bhola P, Tebble C, Shandley K. Preferences for online mental health services among Australian and Indian samples: a cross-cultural comparison. Psychol Stud 2018;63(4):376-383. [doi: 10.1007/s12646-018-0453-y]
26. Schrauben SJ, Appel L, Rivera E, Lora CM, Lash JP, Chen J, CRIC Study Investigators. Mobile health (mHealth) technology: assessment of availability, acceptability, and use in CKD. Am J Kidney Dis 2021;77(6):941-950 [FREE Full text] [doi: 10.1053/j.ajkd.2020.10.013] [Medline: 33309860]
27. Horvath KJ, Bauermeister JA. eHealth literacy and intervention tailoring impacts the acceptability of a HIV/STI testing intervention and sexual decision making among young gay and bisexual men. AIDS Educ Prev 2017;29(1):14-23 [FREE Full text] [doi: 10.1521/aeap.2017.29.1.14] [Medline: 28195779]

28. Steinberg JR, Yeh C, Jackson J, Saber R, Niznik CM, Leziak K, et al. Optimizing engagement in an mHealth intervention for diabetes support during pregnancy: the role of baseline patient health and behavioral characteristics. J Diabetes Sci Technol 2022;16(6):1466-1472 [FREE Full text] [doi: 10.1177/19322968211035441] [Medline: 34423677]
29. Pho AT, Bakken S, Lunn MR, Lubensky ME, Flentje A, Dastur Z, et al. Online health information seeking, health literacy, and human papillomavirus vaccination among transgender and gender-diverse people. J Am Med Inform Assoc 2022;29(2):285-295 [FREE Full text] [doi: 10.1093/jamia/ocab150] [Medline: 34383916]
30. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
31. Maroney K, Curtis LM, Opsasnick L, Smith KD, Eifler MR, Moore A, et al. eHealth literacy and web-based patient portal usage among kidney and liver transplant recipients. Clin Transplant 2021;35(2):e14184. [doi: 10.1111/ctr.14184] [Medline: 33278846]
32. Paige SR, Krieger JL, Stellefson ML. The influence of eHealth literacy on perceived trust in online health communication channels and sources. J Health Commun 2017;22(1):53-65 [FREE Full text] [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
33. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth literacy: patient engagement in identifying strategies to encourage use of patient portals among older adults. Popul Health Manag 2017;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
34. Sarkar M, Sanders LM, Kelleher KJ, Chisolm DJ. Psychosocial health, e-health literacy, and perceptions of e-health as predictors and moderators of e-health use among caregivers of children with special healthcare needs. Telemed J E Health 2016;22(2):123-131 [FREE Full text] [doi: 10.1089/tmj.2015.0028] [Medline: 26295487]
35. Song L, Tatum K, Greene G, Chen RC. eHealth literacy and partner involvement in treatment decision making for men with newly diagnosed localized prostate cancer. Oncol Nurs Forum 2017;44(2):225-233 [FREE Full text] [doi: 10.1011/17.ONF.225-233] [Medline: 28222090]
36. Morton E, Ho K, Barnes SJ, Michalak EE. Digital health literacy in bipolar disorder: international web-based survey. JMIR Ment Health 2021;8(10):e29764 [FREE Full text] [doi: 10.2196/29764] [Medline: 34665143]
37. Bailey CE, Kohler WJ, Makary C, Davis K, Sweet N, Carr M. eHealth literacy in otolaryngology patients. Ann Otol Rhinol Laryngol 2019;128(11):1013-1018. [doi: 10.1177/0003489419856377] [Medline: 31195809]
38. Hogan TP, Hill JN, Locatelli SM, Weaver FM, Thomas FP, Nazi KM, et al. Health information seeking and technology use among veterans with spinal cord injuries and disorders. PM R 2016;8(2):123-130. [doi: 10.1016/j.pmrj.2015.06.443] [Medline: 26164352]
39. Stellefson ML, Shuster JJ, Chaney BH, Paige SR, Alber JM, Chaney JD, et al. Web-based health information seeking and eHealth literacy among patients living with chronic obstructive pulmonary disease (COPD). Health Commun 2018;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
40. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020;39(2):141-150 [FREE Full text] [doi: 10.1177/0733464818807468] [Medline: 30353776]
41. Seçkin G, Hughes S, Yeatts D, Degreve T. Digital pathways to positive health perceptions: does age moderate the relationship between medical satisfaction and positive health perceptions among middle-aged and older internet users? Innov Aging 2019;3(1):igy039 [FREE Full text] [doi: 10.1093/geroni/igy039] [Medline: 30648160]
42. Escoffery C. Gender similarities and differences for e-health behaviors among U.S. adults. Telemed J E Health 2018;24(5):335-343. [doi: 10.1089/tmj.2017.0136] [Medline: 28813630]
43. Park H, Cormier E, Gordon G, Baeg JH. Identifying health consumers' eHealth literacy to decrease disparities in accessing eHealth information. Comput Inform Nurs 2016;34(2):71-76. [doi: 10.1097/CIN.0000000000000205] [Medline: 26657619]
44. Seckin G, Hughes S. Patient-reported outcomes in a nationally representative sample of older internet users: cross-sectional survey. JMIR Aging 2021;4(4):e16006 [FREE Full text] [doi: 10.2196/16006] [Medline: 34822340]
45. Woods SP, Sullivan KL. Lower neurocognitive functioning disrupts the effective use of internet-based health resources in HIV disease: the mediating effects of general health literacy capacity. AIDS Behav 2019;23(3):676-683 [FREE Full text] [doi: 10.1007/s10461-018-2350-8] [Medline: 30506473]
46. Stellefson M, Paige SR, Alber JM, Chaney BH, Chaney D, Apperson A, et al. Association between health literacy, electronic health literacy, disease-specific knowledge, and health-related quality of life among adults with chronic obstructive pulmonary disease: cross-sectional study. J Med Internet Res 2019;21(6):e12165 [FREE Full text] [doi: 10.2196/12165] [Medline: 31172962]
47. Madrigal L, Escoffery C. Electronic health behaviors among US adults with chronic disease: cross-sectional survey. J Med Internet Res 2019;21(3):e11240 [FREE Full text] [doi: 10.2196/11240] [Medline: 30835242]
48. Chung JE, Lee CJ. The impact of cancer information online on cancer fatalism: education and eHealth literacy as moderators. Health Educ Res 2019;34(6):543-555. [doi: 10.1093/her/cyz027] [Medline: 31550372]
49. Schulz PJ, Fitzpatrick MA, Hess A, Sudbury-Riley L, Hartung U. Effects of eHealth literacy on general practitioner consultations: a mediation analysis. J Med Internet Res 2017;19(5):e166 [FREE Full text] [doi: 10.2196/jmir.6317] [Medline: 28512081]

50. Blackstock OJ, Cunningham CO, Haughton LJ, Garner RY, Norwood C, Horvath KJ. Higher eHealth literacy is associated with HIV risk behaviors among HIV-infected women who use the internet. J Assoc Nurses AIDS Care 2016;27(1):102-108 [FREE Full text] [doi: 10.1016/j.jana.2015.09.001] [Medline: 26456347]
51. Aponte J, Nokes KM. Validating an electronic health literacy scale in an older hispanic population. J Clin Nurs 2017;26(17-18):2703-2711. [doi: 10.1111/jocn.13763] [Medline: 28207962]
52. Li X. Understanding eHealth literacy from a privacy perspective: eHealth literacy and digital privacy skills in American disadvantaged communities. Am Behav Sci 2018;62(10):1431-1449. [doi: 10.1177/0002764218787019]
53. Stormacq C, Van den Broucke S, Wosinski J. Does health literacy mediate the relationship between socioeconomic status and health disparities? Integrative review. Health Promot Int 2019;34(5):e1-e17. [doi: 10.1093/heapro/day062] [Medline: 30107564]
54. van der Heide I, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health, education, and health literacy: results from the Dutch adult literacy and life skills survey. J Health Commun 2013;18 Suppl 1(Suppl 1):172-184 [FREE Full text] [doi: 10.1080/10810730.2013.825668] [Medline: 24093354]
55. Jung SO, Son YH, Choi E. E-health literacy in older adults: an evolutionary concept analysis. BMC Med Inform Decis Mak 2022;22(1):28 [FREE Full text] [doi: 10.1186/s12911-022-01761-5] [Medline: 35101005]
56. De Souza RNA, Butt D, Jethani S, Marmo C. Participatory research methods for investigating digital health literacy in culturally and linguistically diverse communities. Transdisciplinary J Cultur Participation 2021;8(1):1-17. [doi: 10.7146/tjcp.v8i1.117800]
57. Xie L, Zhang S, Xin M, Zhu M, Lu W, Mo PKH. Electronic health literacy and health-related outcomes among older adults: a systematic review. Prev Med 2022;157:106997. [doi: 10.1016/j.ypmed.2022.106997] [Medline: 35189203]


## Abbreviations

**COPD:** chronic obstructive pulmonary disease
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses


*Edited by T de Azevedo Cardoso; submitted 01.03.23; peer-reviewed by E Hernandez-Encuentra, V Pérez Jover; comments to author 12.07.23; revised version received 21.07.23; accepted 29.02.24; published 26.07.24.*

*Please cite as:*
*Yuen E, Winter N, Savira F, Huggins CE, Nguyen L, Cooper P, Peeters A, Anderson K, Bhoyroo R, Crowe S, Ugalde A*
*Digital Health Literacy and Its Association With Sociodemographic Characteristics, Health Resource Use, and Health Outcomes: Rapid Review*
*Interact J Med Res 2024;13:e46888*
*URL: https://www.i-jmr.org/2024/1/e46888*
*doi:10.2196/46888*
*PMID:*

Review

# Requirements for and Barriers to Rehabilitation Services for Children With Disabilities in Middle- and High-Income Countries: Scoping Review

Yijun Xie[1], BPA; Jing Wu[2], PhD; Yao Li[1], PhD; Hui Liu[1], BEng; Yanyan Peng[3], BA; Ping Zhou[3], BA; Yizhou Sun[3], BA; Luyan Kang[3], BA; Chenghua Jiang[1], PhD; Hengjing Wu[1], PhD

[1]Clinical Center for Intelligent Rehabilitation Research, Shanghai Yangzhi Rehabilitation Hospital (Shanghai Sunshine Rehabilitation Center), School of Medicine, Tongji University, Shanghai, China

[2]Department of Fundamental Nursing, School of Nursing, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[3]Office of Rehabilitation, Shanghai Disabled Persons' Federation, Shanghai, China

**Corresponding Author:**
Hengjing Wu, PhD
Clinical Center for Intelligent Rehabilitation Research
Shanghai Yangzhi Rehabilitation Hospital (Shanghai Sunshine Rehabilitation Center), School of Medicine
Tongji University
No.2209 Guangxing Road
Songjiang District
Shanghai, 201619
China
Phone: 86 15821525700
Email: whjdata@126.com

## Abstract

**Background:** The rehabilitation of children with disabilities has received considerable attention from the United Nations. However, the state of rehabilitation services for children with disabilities worldwide remains far from optimistic, even in economically affluent middle- and high-income countries.

**Objective:** This scoping review aimed to identify the rehabilitation needs of children with disabilities and their barriers to rehabilitation services in middle- and high-income countries.

**Methods:** A systematic search was conducted using MEDLINE and Web of Science for papers published from January 2013 to December 2023. Studies were included if they were peer-reviewed, full-text articles related to children with disabilities, reporting on their access to rehabilitation services, and conducted in countries classified by the World Bank 2023 as middle- and high-income economies. Exclusion criteria included duplicates, unavailable full texts, and studies without distinct outcomes. A total of 27 studies were selected following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, focusing on children, their families, or service providers.

**Results:** The suitability, availability, and affordability of rehabilitation services were identified as the major needs and barriers for children with disabilities in middle- and high-income countries. This included communication barriers, a need for more personnel and facilities, and the stagnation and inadequacy of economic subsidies.

**Conclusions:** Middle- and high-income countries have relatively well-established rehabilitation infrastructure and support systems. They are nevertheless insufficient for meeting the needs of children with disabilities. More attention should be paid to these issues to improve the well-being of children with disabilities. The data provided by this review can help raise awareness of rehabilitation needs and barriers at the policy level.

***(Interact J Med Res 2024;13:e50047)*** doi:10.2196/50047

**KEYWORDS**

children with disabilities; barriers; health services; middle- and high-income countries; child; low income; middle income; disability; children; disabilities; income; barrier; rehabilitation; suitability; availability; affordability; support system; support; awareness; policy

## Introduction

Worldwide, over 240 million children are impacted by disabilities [1], which predominantly manifest as functional impairments. These impairments substantially inhibit their ability to engage in foundational activities such as learning, daily functions, and social integration. Defined by age, individuals younger than 18 years who experience physiological, psychological, or cognitive deficits, whether congenital or acquired, are classified as children with disabilities [2]. Children are in a crucial period of growth and development. Timely intervention and treatment not only promote the physical and cognitive development of children with disabilities but also minimize the impact of disabilities on their lives to the greatest extent possible [3]. Rehabilitation for children with disabilities encompasses a comprehensive approach that integrates medical, educational, and social rehabilitation methods. Its primary aim is to assist children in overcoming physiological, psychological, or societal barriers. The ultimate goal is to maximize their quality of life, functionality, and autonomy [4]. Compared with their healthy peers, children with disabilities have higher rates of morbidity, require increased medical attention, and have longer hospital stays [5]. The rehabilitation needs of children with disabilities are more pressing. However, children with disabilities often face greater challenges in meeting their rehabilitation needs. Children with disabilities sometimes face the impact of different social attitudes and cultural traditions, experiencing discrimination and exclusion, which results in their service needs being overlooked [5]. Furthermore, barriers related to transportation and the natural environment also limit the fulfillment of the needs [6].

Rehabilitation welfare policies for children with disabilities have been established in most regions [7-9]. However, these policies seem to have not yielded satisfactory outcomes [10]. Children with disabilities still face considerable drawbacks in various indicators used to assess child welfare. In sub-Saharan African countries [11], common deficiencies, such as attitudinal problems, poverty, inadequately trained, health care professionals, and physical inaccessibility are seen. The existence of regional disparities contributes to the insufficient distribution of health care resources and funding [12,13]. This further affects the accessibility of rehabilitation services for children with disabilities residing in different regions. The disparity in national economic levels directly impacts the allocation of medical resources and funding, resulting in significant variations in the level of support available for children with disabilities residing in different regions [14,15]. The World Bank classifies the global economy into 4 distinct categories based on 3 income thresholds: US $1135, US $4465, and US $13,846 per capita gross national income (GNI) for the year 2022. These categories are low-income, lower-middle-income, upper-middle-income, and high-income [16]. Previous research has predominantly focused on countries classified as low-income economies [12,17,18]. Even in middle- and high-income countries with relatively better economic conditions, the status of rehabilitation support for children with disabilities is equally concerning [19]. Compared with low-income countries, these nations and regions possess a greater abundance of resources and experience in providing rehabilitation services. However, there are still issues with the accessibility of rehabilitation services for children with disabilities in these regions [17]. The growth of the population and the increasing prevalence of chronic illnesses have raised additional demands for rehabilitation services. Disparities in economic levels have exacerbated regional inequalities in the allocation of rehabilitation resources [18]. Additionally, differences between early-established welfare policies and current rehabilitation needs also impact the fulfillment of these demands [20]. There has been relatively limited exploration in existing literature regarding the specific influencing factors that pertain to middle- to high-income countries and regions. The purpose of this study was to systematically review the existing literature on the state of rehabilitation services for children with disabilities, especially in middle- and high-income countries. This review aims to explore the factors influencing the accessibility and effectiveness of these services and to identify gaps in current policies that could be addressed to better meet the diverse needs of children with disabilities across different economic contexts.

## Methods

### Review Design

We conducted a scoping review to map the evidence regarding the accessibility of rehabilitation services for children with disabilities across the world. The review was based on a systematic search of relevant sources that have been used previously in the field of disability [21,22]. Scoping reviews have broad, comprehensive objectives compared with systematic reviews, which are often guided by more narrow, focused research questions [23]. We chose this approach given the limited range of rigorous study designs described in traditional systematic review papers. In addition, because this scoping review was intended to provide a descriptive overview of the applicable literature [24], we did not critically assess the papers. We acknowledge that the studies included in the review may have methodological strengths and limitations. Therefore, we approach the selection and analysis with caution. We are particularly mindful that some studies may not be methodologically suited to provide precise explanations. However, we make an effort to highlight the contributions of each study and their role in the broader context. While scoping reviews may not be the most robust tool for evidence synthesis, they provide us with a valuable starting point to understand the unmet health care needs of children with disabilities and point the way for future research. The outline for this review followed the 5-stage framework described by Arksey and O'Malley [23], which includes (1) identifying the research question, (2) identifying relevant studies, (3) selecting studies, (4) charting the data, and (5) accumulating, summarizing, and reporting the results.

### Identifying the Research Question

The purpose of this scoping review was to understand the current status of rehabilitation services for children with disabilities, explore the influencing factors, and make constructive observations. This is important for improving the well-being of

children with disabilities and responding to the World Health Organization's (WHO's) 2030 Rehabilitation Initiative.

### Identifying Relevant Studies

We searched the electronic databases MEDLINE and Web of Science using combinations of the keywords, which are Rehabilitation, Disabled Children, and Health Services for Persons with Disabilities. Specific search strategies can be found in Multimedia Appendix 1. We screened the titles and abstracts of publications from the past 10 years (2013-2023) to find relevant articles and capture the most recent evidence. The literature within this time frame encompasses alterations in policy, technology, and rehabilitation services, thereby augmenting the filterability and analyzability of research data and literature [25-27]. Through this strategy, we aimed to provide highly relevant and practical information.

### Eligibility Criteria and Study Selection

Studies were eligible if they met the following criteria: (1) a study related to a child with a disability; (2) the participants were children with disabilities, their family members (parents, relatives), or staff providing services; (3) the study reported the disabled child's access to rehabilitation services; and (4) the study was conducted in countries classified by the World Bank 2023 as middle- and high-income economies. Studies were excluded if the full text of the electronic source was not available. Duplicate reports of the same study were combined if they reported different results or excluded if they had the same results. The first round of screening removed duplicate studies and eliminated articles based on titles and abstracts. Our selection process was conducted in 2 stages to ensure accuracy and relevance. In the first stage, XYJ and LY independently reviewed titles, abstracts, and full texts of identified publications. Each conducted their assessments separately to maintain objectivity, documenting their findings to ensure alignment with our inclusion criteria.

In the second stage, these authors met to jointly reassess the retained papers, focusing specifically on their relevance to our review's topic. This collaborative review helped refine the selection further. Any disagreements encountered during this reassessment were resolved through detailed discussions until a consensus was reached regarding each paper's compliance with the inclusion criteria.

### Data Charting

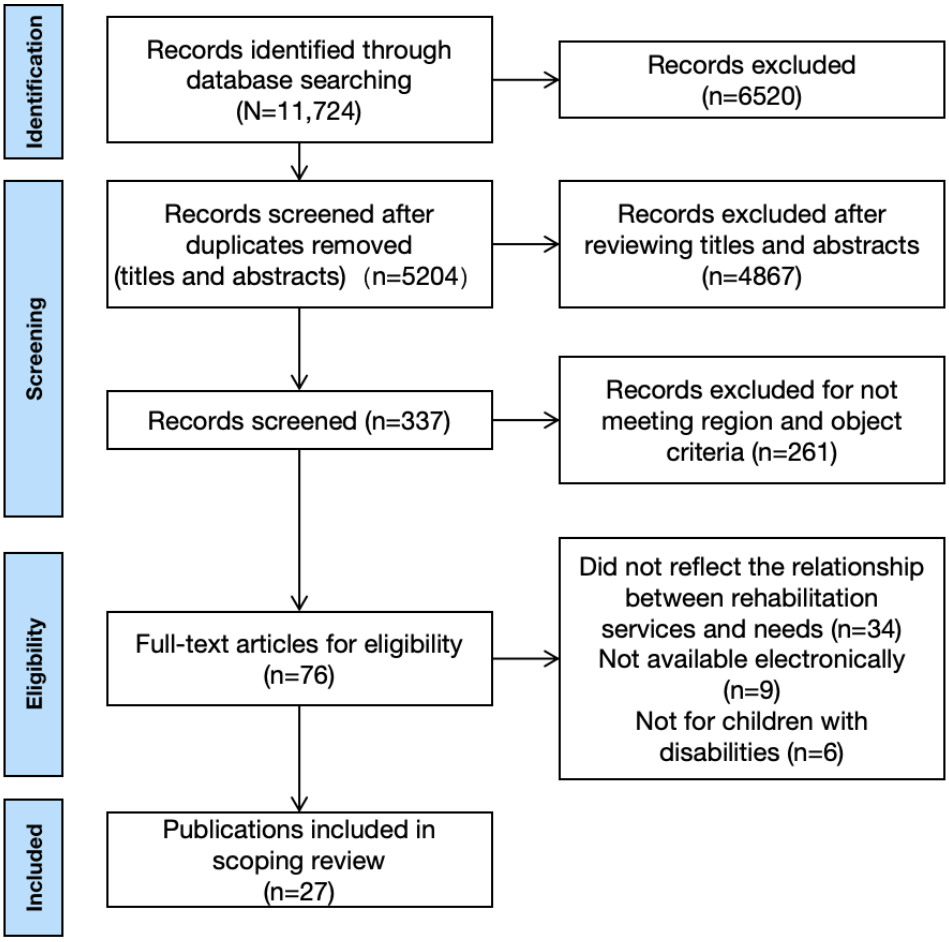
Participant demographics (eg, information on children with disabilities and type of disability), access to rehabilitation services, and study-related data (eg, the author or researcher, year of publication, study objectives, and whether the study was in a middle- or high-income country) were extracted and put into Excel (Microsoft Corp). Two independent reviewers conducted the extraction of pertinent data from the papers included in the scope-defining review. Next, we read the full texts of the remaining papers and extracted the following: the author or researcher, date of publication, study objectives, inclusion criteria, age of child or children, sample size, age of parents, type of disability, city or country or setting of the study, research method, study results, and the facilitators of and barriers to rehabilitation services. Any disagreements among reviewers that cannot be resolved through discussion or consensus will be resolved by a third reviewer. Penchansky and Thomas [28] have compiled a set of specific dimensions that delineate the relationship between patients and the health care system. Their analysis demonstrates discernible distinctions among these dimensions. These specific dimensions encompass the aspects of availability, accessibility, accommodation, affordability, and acceptability. The extracted information has been categorized using these 5 dimensions.

## Results

### Search Results

The initial search produced 11,724 documents after removing 6520 duplicates; 4867 documents were further removed after screening titles and abstracts. We then screened 76 full-text papers and excluded an additional 49 papers. The reasons for exclusion during full-text screening were the following: (1) did not reflect the relationship between rehabilitation services and need (n=34); (2) full text was not available electronically (n=9); and (3) did not target children with disabilities (n=6). Figure 1 illustrates the search process. All the included studies are listed in Multimedia Appendix 2 [15,29-54].

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram of the study selection process.




## Characteristics of the Included Articles

The majority of the studies (21/27, 77.8%) of children's rehabilitation services were conducted in high-income economies, with 22.2% (6/27) at upper-middle income levels. Regarding regions, the highest number of studies were conducted in East Asia and Pacific (9/27, 33.4%) and Europe and Central Asia (8/27, 29.6%), followed by North America (7/27, 25.9%), Middle East and North Africa (2/27, 7.4%), and sub-Saharan Africa (1/27, 3.7%).

The selected studies investigated the rehabilitation needs of children with disabilities and the barriers they face. Of these studies, 33.4% (9/27) solely involved children with disabilities, while the remaining papers gathered information from parents of children with disabilities (12/27, 44.4%) or professional caregivers (4/27, 14.8%). In some studies, children with disabilities could not articulate their needs due to their unique physical conditions. Consequently, researchers gained insights into the children's rehabilitation requirements and challenges by conducting interviews with their parents. For instance, parents were asked about their perceptions of their child's experience with rehabilitation services or to elaborate on the effect of their child's disability on the family's quality of life.

Overall, the included studies had diverse goals, such as investigating the unmet health care needs of children with disabilities, insurance plans, quality of life for children and families, participation in activities, and intervention programs. Nevertheless, the primary focus of the reviewed papers was on the health care requirements of children with disabilities and the barriers they face, with relatively limited attention given to treatment techniques or clinical outcomes. Table 1 lists the descriptive statistics of the included documents. All selected documents will be included as supplementary materials in the form of appendices.

**Table 1.** General characteristics of the studies eligible for inclusion in the review.

| Characteristics | Publications |
|---|---|
| **World Bank income category, n (%)** | |
| High-income economy | 21 (77.8) |
| Upper-middle income economy | 6 (22.2) |
| **World Bank region, n (%)** | |
| East Asia and Pacific | 9 (33.4) |
| Europe and Central Asia | 8 (29.6) |
| North America | 7 (25.9) |
| Middle East and North Africa | 2 (7.4) |
| Sub-Saharan Africa | 1 (3.7) |
| **Participants, n (%)** | |
| Parent | 12 (44.4) |
| Children | 9 (33.4) |
| Caregiver | 4 (14.8) |
| Multiple | 2 (7.4) |
| **Disability domain, n (%)** | |
| Multiple domains | 17 (62.9) |
| Physical | 5 (18.5) |
| Intellectual | 2 (7.4) |
| Neurological | 2 (7.4) |
| Hearing | 1 (3.7) |
| **Study design, n (%)** | |
| Quantitative | 16 (59.3) |
| Qualitative | 10 (37) |
| Mixed methods | 1 (3.7) |

## Rehabilitation Needs and Barriers to Accessing Services

The scale by Penchansky and Thomas [28] addresses access to health care and consists of 5 categories, which include accessibility, acceptability, affordability, appropriateness, and availability. This classification's validity has been confirmed through interview data analysis. Accessibility describes the physical connection between the family and the location of rehabilitation services. Acceptability refers to clients' and providers' satisfaction with health care services. Affordability relates to the connections between financial constraints, financial capacity, and health insurance availability. Appropriateness refers to the match between supply and patient demand. Availability describes the connection between supply and demand for resources. This study used Penchansky's scale to identify information about barriers to rehabilitation services (Table 2).

**Table 2.** Factors affecting the fulfillment of rehabilitation needs of children with disabilities.

| Factors | Reference |
|---|---|
| **Accessibility** | |
| Transportation problems | Jeong [40] |
| Distance to service | Alyami et al [29], Mulligan et al [41], Caicedo [36] |
| **Acceptability** | |
| Rejection and discrimination | Wang et al [39] |
| Privacy and trust | Teleman et al [42] |
| Shame and fear | Mulligan et al [41], Teleman et al [42], Wang et al [39] |
| Doubts about validity | Teleman et al [42], Sukeri et al [37], Lindly et al [43], Xia et al [44] |
| **Affordability** | |
| Insurance issues | Lindly et al [43], Schaible et al [30], Robinson et al [33], Xia et al [44] |
| Burden of treatment | Alyami et al [29], Lindly et al [43], Parr et al [32], Roux-Levy et al [38], Piškur et al [45], Xia et al [44], Houtrow et al [46], Raouafi et al [15] |
| Other economic issues | Lindly et al [43], Schaible et al [30], Robinson et al [33], Gallagher et al [34], Raouafi et al [15], Jeong [40], Meehan et al [47], Umat et al [48] |
| **Appropriateness** | |
| Number of treatments | Alyami et al [29], He et al [49] |
| Targeted rehabilitation services | Sukeri et al [37], Wang et al [39] |
| Expertise | Mulligan et al [41], Sukeri et al [37], Roux-Levy et al [38], Wang et al [39] |
| Communication barriers | Carter et al [50], Mulligan et al [41], Sukeri et al [37], Parr et al [32], Caicedo [36], Ziviani et al [35], Pérez-Ardanaz et al [51] |
| **Availability** | |
| Procedures | Roux-Levy et al [38], Sukeri et al [37] |
| Schedule | Wang et al [39], Alyami et al [29], Ziviani et al [35] |
| Messaging | Xia et al [44], Alyami et al [29], Matsuzawa and Shiroki [52], Ziviani et al [35], Sukeri et al [37] |
| Spiritual support | Xia et al [44], Alyami et al [29], Matsuzawa and Shiroki [52], Mulligan et al [41], Piškur et al [45], Cacioppo et al [31], Lindly et al [43], Khusaifan and Keshky [53], Umat et al [48] |
| Equipment service supply | Carter et al [50], Pérez-Ardanaz et al [51], Alyami et al [29], Mulligan et al [41], Houtrow et al [46], Roux-Levy et al [38], Arabiat et al [54], Jeong [40], Sukeri et al [37] |

### *Accessibility*

The accessibility of rehabilitation services refers to the reduction of various barriers that may exist in accessing these services, intending to achieve equality for all. Ensuring that children with disabilities can easily access rehabilitation services has a significant impact on meeting the rehabilitation needs of these children [55]. On the one hand, geographic distance is strongly linked with time costs for children with disabilities to access rehabilitation [29]. On the other hand, the distance barrier also interferes with the ability of rehabilitation specialists to connect and share experiences. Geographical isolation hinders information exchange in resource-poor regions, often leaving rehabilitation experts in these areas isolated and unsupported "information islands" [27]. Apart from the challenges associated with the coverage and frequency of "public transportation," Jeong [40] found that "personal transportation" includes factors such as the availability of vehicles also influencing the accessibility of rehabilitation services.

### *Acceptability*

The acceptance of rehabilitation services by children with disabilities is affected by personal factors such as stigmatization and doubts about service effectiveness [28,29,55,56]. According to 3 papers [39,41,42], disability is often stigmatized, leading to a reluctance among parents to seek rehabilitation treatment for their children with disabilities. This stigma associated with disability can deter families from pursuing the necessary care and therapies [57]. Discrimination may also lead to rehabilitation professionals refusing to provide their services. This directly impacts the trust relationship between children with disabilities and service providers, ultimately influencing the decision to seek rehabilitation services [30]. In a reviewed study [58], a Chinese mother faced discrimination due to her child's autism, leading to trust issues in rehabilitation services and service discontinuation. The current situation of this discrimination may explain how children with disabilities are deprived of their rehabilitation rights through voluntary abandonment in a so-called equitable institutional environment. Furthermore,

rehabilitation care is a long-term process. Short-term outcomes are often limited. Some families of children with disabilities may develop a negative attitude toward rehabilitation services because they perceive that the short-term impact falls short of their initial expectations [56].

### *Affordability*

The relationship between disability and poverty is reciprocal [56,57]. Families of children with disabilities were more likely to experience financial burdens owing to their child's health condition [30]. Such families face multiple financial challenges, including rehabilitation expenses [30,58,59], assistive equipment, institutional care [29], home renovations [15], and travel expenses [60]. While governments in middle- to high-income countries generally offer financial assistance and insurance, some families having children with disabilities still experience significant economic burdens [61-63], and current insurance benefits fall short of completely covering the rehabilitation needs of children with disabilities [31]. The gap between financial assistance and the actual needs continues to widen [32,33]. It is worth noting that the current financial assistance system primarily focuses on children with confirmed disabilities. This might exclude children with disabilities who require rehabilitation but do not have a formal diagnosis [34].

### *Appropriateness*

Effective rehabilitation treatment relies heavily on the appropriateness of rehabilitation services, which includes matching treatment services, goals, and assistive devices with patients' needs [64]. As reported by mothers in a study conducted in Kelantan, Malaysia, profit-oriented organizations may overlook the individual needs of children with disabilities, which can diminish the effectiveness of rehabilitation services. Parents have noted that insufficient responsiveness from professionals has reduced the practical utility of the services [56]. The professionalism of rehabilitation professionals is crucial for the appropriateness of services. Insufficient professionalism can result in delays in addressing the conditions of children with disabilities. Rehabilitation training is widely available in middle- to high-income countries. However, there is a lack of professional ability among rehabilitation personnel. This deficiency leads to children with disabilities not receiving timely and accurate treatment, even when they diligently follow their rehabilitation plans [56]. Due to limitations in the size and staff of rehabilitation institutions, the frequency and number of treatments can also impact the effectiveness and satisfaction of rehabilitation services. In a family survey of Saudi Arabian children with hearing impairment [26], more than half of the participants believed that the number of treatment sessions provided was inadequate to meet their rehabilitation demands.

Communication gaps between the staff of rehabilitation institutions and the parents of children with disabilities can lead to misunderstandings and negative perceptions [35,36]. A lack of collaborative communication among staff within rehabilitation institutions can increase feelings of isolation among rehabilitation professionals. It also has the potential to reduce job motivation [28]. Even though rehabilitation facilities in middle- and high-income countries are relatively well-established, the lack of communication and interaction has still been identified as a serious issue. Studies indicate that poor internal and external communication within institutions, along with a lack of collaboration, increases the co-ordination burden on parents of children with disabilities [58]. This also hinders the transfer of treatment between different rehabilitation systems and the sustainable development of children with disabilities [32]. The compatibility of assistive facilities is also a key factor. For instance, in the provision of rehabilitation equipment for children with physical disabilities, nonportable wheelchairs are often voluntarily abandoned because they do not meet the actual needs of children with disabilities [37].

### *Availability*

The availability of rehabilitation services for children with disabilities emphasizes the presence and provision of rehabilitation resources, as well as the reasonable scheduling of rehabilitation service delivery, which are key factors in meeting the rehabilitation needs of children with disabilities. Factors related to this include cumbersome procedures, the need for information and emotional support, the need for specialized medical facilities, and scheduling convenience. If the family cannot afford the cost of the wait times, they will give up on rehabilitation [38]. Australian parents of children with physical disabilities report that flexible scheduling and location of rehabilitation services provide them with convenience [35].

Studies highlight the need for information on rehabilitative options [29]. Available research showed that middle- and high-income countries have relatively well-established rehabilitation facility structures, yet the exchange and communication of rehabilitation information have not received adequate attention. The quantity and quality of information provided by rehabilitation service institutions are often unsatisfactory. This compels parents of children with disabilities to seek solutions on their own [58]. It is worth noting that children with disabilities and their caregivers experience both physical and mental stress. While it has been proven that psychological support is beneficial for their mental well-being [26,40,41], it remains one of the common unmet needs within rehabilitation programs [42,57]. Regional disparities in rehabilitation services create challenges for children with disabilities who seek higher-quality rehabilitation services [38]. To access higher-level services, they are compelled to bear higher treatment costs [29,39,65]. The current situation of oversaturated health care resources and low-quality health system services also prolongs the access cycle and reduces the level of services for children with disabilities [38,55].

The most commonly discussed issue in the literature was availability, followed by affordability. Notably, Europe or Central Asia, North America, and East Asia or Pacific were the main regions reporting both of these issues.

High-income countries reported the greatest need for specialized facilities and emotional support, followed by financial issues not covered by insurance and treatment costs. Communication barriers and doubts about the effectiveness of rehabilitation treatment were also highlighted as acceptability concerns. Professionalism and access to spiritual support were also mentioned. In contrast, participants from middle-income countries expressed greater concern with the accessibility,

appropriateness, and acceptability of rehabilitation services. This disparity might be related to the fact that half of the studies in middle-income countries focused on rehabilitation service workers.

## Discussion

### Principal Findings

This scoping review aimed to provide an overview of the rehabilitation needs and related obstacles faced by children with disabilities residing in middle- and high-income countries. The final review comprised 27 peer-reviewed papers published between 2013 and 2023, categorized into 5 groups based on rehabilitation-related factors, allowing for a summary and comparison of related studies. The findings of this review suggest that even in economically prosperous middle- and high-income countries, common factors still affect children with disabilities' achievement of rehabilitation services. The appropriateness, accessibility, and affordability of rehabilitation services were identified as the main factors related to the needs and barriers of families of children with disabilities. Having identified these factors, it is possible to propose recommendations aimed at the formulation of intervention strategies.

The accessibility of rehabilitation services is a fundamental issue related to the presence or absence of service resources. In line with research conducted in low-income countries, it is consistent that children with disabilities generally face difficulties in accessing adequate health care services [66]. This issue is not only confined to low-income countries but also in middle- and high-income countries with more developed health care systems [67]. However, there are differences between the 2 when it comes to the specific challenges they face: unlike the generally low availability of rehabilitation resources in low-income countries, the main influence factor in middle- and high-income countries is the uneven distribution. The availability of rehabilitation resources is generally high but concentrated in economically developed areas [68]. In underdeveloped regions, children with disabilities face challenges in accessing rehabilitation resources, resulting in a relative scarcity of rehabilitation resource provision [28]. Capitalizing on the spillover effect of developed regions, rehabilitation resources should be strategically allocated to less developed regions. This can improve the overall quality of services for families of children with disabilities, enhancing the capabilities of existing health care and rehabilitation professionals while attracting new talent to underserved areas. By increasing both the quantity and quality of professionals, the accessibility and effectiveness of services could be improved. Through the implementation of these measures, we can mitigate the uneven distribution of resources between regions and reduce disparities in rehabilitation levels [68].

Access to affordable and adequate health care for children with disabilities remains a pervasive challenge in middle- and high-income regions. It was discussed in 12 papers in 8 different countries. In contrast to low-income countries, the economic pressure in this context does not arise from a lack of assistance but rather from the effectiveness of the assistance system itself [26,35,64]. As economic levels and rehabilitation costs continue to rise, the actual medical expenses surpass the financial support available to families. This discrepancy has rendered the previously established financial subsidy policies inadequate in meeting the current rehabilitation support needs. This diminishes the effectiveness of assistance. Although such countries generally have relatively completed insurance systems and benefit arrangements [25], an unmet need for support also exists. For instance, current insurance benefits fall short of completely covering rehabilitation costs for children with disabilities [39]. Evaluating the effectiveness of financial assistance programs and adjusting subsidy standards for children with disabilities based on actual circumstances is necessary. By understanding the specific needs of different types of children with disabilities and making targeted adjustments to subsidy policies, resource allocation can be improved, leading to enhanced utilization efficiency. Extending insurance coverage to encompass a diverse range of rehabilitation-related expenditures will better meet the individualized needs of children with disabilities. Besides traditional physical and speech therapy, occupational therapy, psychological counseling, and special education resources should also be incorporated. Achieving this goal necessitates close co-operation among health care service providers, policymakers, and other stakeholders to adopt policies to address complex rehabilitation needs. Simultaneously, the long-term and intricate nature of the rehabilitation process for children with disabilities must be considered, and adjustments in assistance to offer more flexible and enduring support should be made.

Effective communication is a pivotal issue in improving the appropriateness of rehabilitation for children with disabilities. In low-income countries, this issue has received relatively little attention in research [12]. However, in this review, 6 out of 15 middle- to high-income countries (40%) have reported this concern. This encompasses the need for communication between children with disabilities and rehabilitation staff, knowledge exchange among rehabilitation professionals, and referrals between rehabilitation institutions. Educating and training rehabilitation service providers in effective communication strategies is crucial, encouraging them to actively listen to families' opinions and address their concerns. Future research should delve into the collaboration between rehabilitation experts and institutions, emphasizing resource sharing, information flow, and referral and treatment models for children with disabilities. This approach aims to enhance the sustainability of rehabilitation services for children with disabilities.

This review has several limitations that should be noted. First, even though we used an extensive search strategy to find all relevant studies, the search scope was constrained because our study was not a systematic review. Second, most studies were conducted in North America, Europe, and Asia or the Pacific, making comparisons by continent and economic development level difficult. Furthermore, some of the studies were carried out at the district level, making it difficult to extrapolate to the nation's overall rehabilitation status. Finally, rather than focusing on more general disability identification, interventions, and clinical research, we mainly considered the rehabilitation

needs of children with disabilities and the related barriers. Therefore, this might not reflect access to rehabilitation services in the broadest sense, as mentioned in WHO's 2030 Rehabilitation Initiative. Future research could further refine rehabilitation initiatives related to children with disabilities to obtain more complete information.

## Conclusion

This scoping review highlighted the rehabilitation needs of children with disabilities and the barriers to rehabilitation services in middle- and high-income countries. The affordability, availability, and appropriateness of rehabilitation services were major factors affecting access. Despite the established service supply in these countries, access falls short of meeting the needs of children with disabilities. To address these challenges, it is necessary to expand financial coverage, improve infrastructure and professional training, and strengthen supporting institutions.

## Acknowledgments

We gratefully acknowledge the support and resources provided by the Department of Public Health and Preventive Medicine, Shanghai Tongji University Medical College, and Shanghai Yangzhi Rehabilitation Hospital (Shanghai Sunshine Rehabilitation Center). Their contributions were vital to the success of this research. The study was funded through a Health Management Outstanding Talents of Shanghai Association of Rehabilitation Medical (2023JGYQ08), Clinical Research Plan of SHDC (SHDC2022CRS049), Rehabilitation Policy Project of Shanghai Disabled Persons' Federation (2022ZC1001 and 2023ZC1001), and Shanghai Hospital Development Center Foundation—Shanghai Municipal Hospital Rehabilitation Medicine Specialty Alliance (SHDC22023304).

## Data Availability

All data generated and analyzed during this study are included in this published paper and its Multimedia Appendices.

## Authors' Contributions

XYJ and LY performed data curation, writing, reviewing, and editing of the original draft. WJ and LH contributed to the investigation and editing. WHJ performed supervision and project administration. All the authors have read and agreed to the published version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Retrieval strategy.
[PDF File (Adobe PDF File), 80 KB - ijmr_v13i1e50047_app1.pdf ]

Multimedia Appendix 2
Included literature.
[PDF File (Adobe PDF File), 99 KB - ijmr_v13i1e50047_app2.pdf ]

Multimedia Appendix 3
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 131 KB - ijmr_v13i1e50047_app3.pdf ]

## References

1. Nearly 240 million disabled children around the world face multiple challenges in realizing their rights. United Nations News. 2021 Nov 10. URL: https://news.un.org/zh/story/2021/11/1094192 [accessed 2023-03-05]
2. International Classification of Functioning, Disability and Health: Children & Youth Version: ICF-CY. Geneva, Switzerland: World Health Organization; 2007.
3. Bickenbach J. The world report on disability. Disability & Society 2011;26(5):655-658. [doi: 10.1080/09687599.2011.589198]
4. Rehabilitation in health systems. Geneva, Switzerland: World Health Organization; 2017.
5. Tilahun D, Hanlon C, Fekadu A, Tekola B, Baheretibeb Y, Hoekstra RA. Stigma, explanatory models and unmet needs of caregivers of children with developmental disorders in a low-income African country: a cross-sectional facility-based survey. BMC health services research 2016 Apr 27;16:152 [FREE Full text] [doi: 10.1186/s12913-016-1383-9] [Medline: 27117326]

6. Nota A, Chikwanha TM, January J, Dangarembizi N. Factors contributing to defaulting scheduled therapy sessions by caregivers of children with congenital disabilities. Malawi Medical Journal 2015;27(1):25-28 [FREE Full text] [doi: 10.4314/mmj.v27i1.7] [Medline: 26137195]
7. Nuri RP, Aldersey HM, Ghahari S, Huque AS. Experience of families in accessing government-led support for children with disabilities in Bangladesh. Disability and Rehabilitation 2022;44(8):1354-1366. [doi: 10.1080/09638288.2020.1804000] [Medline: 32809849]
8. Gaskin DJ, Mitchell JM. Health status and access to care for children with special health care needs. The Journal of Mental Health Policy and Economics 2005;8(1):29-35. [Medline: 15870483]
9. da Cunha MAO, Santos HF, de Carvalho MEL, Miranda GMD, de Albuquerque MDSV, de Oliveira RS, et al. Health care for people with disabilities in the unified health system in Brazil: a scoping review. Int J Environ Res Public Health 2022;19(3):1472 [FREE Full text] [doi: 10.3390/ijerph19031472] [Medline: 35162497]
10. Dempsey I, Keen D. A review of processes and outcomes in family-centered services for children with a disability. Topics in Early Childhood Special Education 2008;28(1):42-52. [doi: 10.1177/0271121408316699]
11. Adugna MB, Nabbouh F, Shehata S, Ghahari S. Barriers and facilitators to healthcare access for children with disabilities in low and middle income sub-Saharan African countries: a scoping review. BMC health services research 2020;20(1):15 [FREE Full text] [doi: 10.1186/s12913-019-4822-6] [Medline: 31906927]
12. Birenbaum A, Guyot D, Cohen HJ. Health care financing for severe developmental disabilities. Monographs of the American Association on Mental Retardation 1990(14):1-150. [Medline: 1703621]
13. Szilagyi PG. Managed care for children: effect on access to care and utilization of health services. Future Child 1998;8(2):39-59. [Medline: 9782649]
14. Fox HB, McManus MA. Improving state Medicaid contracts and plan practices for children with special needs. Future Child 1998;8(2):105-118. [Medline: 9782653]
15. Raouafi S, Achiche S, Raison M. Socioeconomic disparities and difficulties to access to healthcare services among Canadian children with neurodevelopmental disorders and disabilities. Epidemiol Health 2018;40:e2018010 [FREE Full text] [doi: 10.4178/epih.e2018010] [Medline: 29642656]
16. World bank country and lending groups. World Bank Data. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2023-03-06]
17. Graaf G, Gigli K. Care coordination and unmet need for specialised health services among children with special healthcare needs in the USA: results from a cross-sectional analysis of the national survey of children with special healthcare needs. BMJ Open 2022;12(11):e063373 [FREE Full text] [doi: 10.1136/bmjopen-2022-063373] [Medline: 36385031]
18. Parish SL, Shattuck PT, Rose RA. Financial burden of raising CSHCN: association with state policy choices. Pediatrics 2009;124 Suppl 4:S435-S442. [doi: 10.1542/peds.2009-1255P] [Medline: 19948610]
19. Kogan MD, Newacheck PW, Blumberg SJ, Heyman KM, Strickland BB, Singh GK, et al. State variation in underinsurance among children with special health care needs in the United States. Pediatrics 2010;125(4):673-680. [doi: 10.1542/peds.2009-1055] [Medline: 20211947]
20. Horridge KA, Dew R, Chatelin A, Seal A, Macias LM, Cioni G, European Academy of Childhood Disability. Austerity and families with disabled children: a European survey. Developmental Medicine & Child Neurology 2019;61(3):329-336 [FREE Full text] [doi: 10.1111/dmcn.13978] [Medline: 30028502]
21. McPherson AC, Keith R, Swift JA. Obesity prevention for children with physical disabilities: a scoping review of physical activity and nutrition interventions. Disability and Rehabilitation 2014;36(19):1573-1587. [doi: 10.3109/09638288.2013.863391] [Medline: 24308905]
22. Hunt X, Watermeyer B. A web of gaps: a discussion of research strands concerning global south families with a disabled child. Global Health Action 2017;10(1):1337355 [FREE Full text] [doi: 10.1080/16549716.2017.1337355] [Medline: 28649932]
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
24. Cleaver S, Nixon S. A scoping review of 10 years of published literature on community-based rehabilitation. Disability and Rehabilitation 2014;36(17):1385-1394. [doi: 10.3109/09638288.2013.845257] [Medline: 24151820]
25. The state of the world's children 2013: children with disabilities: from exclusion to inclusion. UNICEF. URL: https://www.unicef.org/reports/state-worlds-children-2013 [accessed 2024-05-14]
26. Olusanya BO, Nair MKC, Wright SM, Hadders-Algra M, Global Research on Developmental Disabilities Collaborators (GRDDC). Global progress towards early childhood development for children with disabilities, 2013-2023. BMJ Paediatrics Open 2023;7(1):e002046 [FREE Full text] [doi: 10.1136/bmjpo-2023-002046] [Medline: 37460256]
27. Early Childhood Development and Disability: A Discussion Paper. Geneva, Switzerland: World Health Organization; 2012.
28. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. Medical Care 1981;19(2):127-140. [doi: 10.1097/00005650-198102000-00001] [Medline: 7206846]
29. Alyami H, Soer M, Swanepoel A, Pottas L. Deaf or hard of hearing children in Saudi Arabia: status of early intervention services. International Journal of Pediatric Otorhinolaryngology 2016;86:142-149. [doi: 10.1016/j.ijporl.2016.04.010] [Medline: 27260597]

30. Schaible B, Colquitt G, Caciula MC, Carnes A, Li L, Moreau N. Comparing impact on the family and insurance coverage in children with cerebral palsy and children with another special healthcare need. Child: Care, Health and Development 2018;44(3):370-377. [doi: 10.1111/cch.12547] [Medline: 29327378]
31. Cacioppo M, Bouvier S, Bailly R, Houx L, Lempereur M, Mensah-Gourmel J, ECHO Group. Emerging health challenges for children with physical disabilities and their parents during the Covid-19 pandemic: the ECHO French survey. Annals of Physical and Rehabilitation Medicine 2021;64(3):101429 [FREE Full text] [doi: 10.1016/j.rehab.2020.08.001] [Medline: 32818674]
32. Parr JR, Jolleff N, Gray L, Gibbs J, Williams J, McConachie H. Twenty years of research shows UK child development team provision still varies widely for children with disability. Child: Care, Health and Development 2013;39(6):903-907. [doi: 10.1111/cch.12025] [Medline: 23425219]
33. Robinson LR, McCord RF, Cloud LK, Kaminski JW, Cook A, Amoroso J, et al. Trends over time and jurisdiction variability in supplemental security income and state supplementary payment programs for children with disabilities. Journal of public health management and practice 2020;26 Suppl 2:S45-S53 [FREE Full text] [doi: 10.1097/PHH.0000000000001122] [Medline: 32004222]
34. Gallagher AL, Galvin R, Robinson K, Murphy CA, Conway PF, Perry A. The characteristics, life circumstances and self-concept of 13 year olds with and without disabilities in Ireland: a secondary analysis of the growing up in Ireland (GUI) study. PLoS One 2020;15(3):e0229599 [FREE Full text] [doi: 10.1371/journal.pone.0229599] [Medline: 32168358]
35. Ziviani J, Darlington Y, Feeney R, Rodger S, Watter P. Early intervention services of children with physical disabilities: complexity of child and family needs. Australian Occupational Therapy Journal 2014;61(2):67-75. [doi: 10.1111/1440-1630.12059] [Medline: 24689917]
36. Caicedo C. Families with special needs children: family health, functioning, and care burden. Journal of the American Psychiatric Nurses Association 2014;20(6):398-407. [doi: 10.1177/1078390314561326] [Medline: 25428686]
37. Sukeri S, Bakar RS, Othman A, Ibrahim MI. Barriers to unmet needs among mothers of children with disabilities in Kelantan, Malaysia: a qualitative study. Journal of Taibah University Medical Sciences 2017;12(5):424-429 [FREE Full text] [doi: 10.1016/j.jtumed.2017.05.002] [Medline: 31435274]
38. Roux-Levy PH, Sanlaville D, De Freminville B, Touraine R, Masurel A, Gueneau I, et al. Care management in a French cohort with down syndrome from the AnDDI-Rares/CNSA study. European Journal of Medical Genetics 2021;64(10):104290. [doi: 10.1016/j.ejmg.2021.104290] [Medline: 34274527]
39. Wang Y, Zhu Y, Qi CY, Zhang Q. An exploration of voluntarily abandoned free health services among children with disabilities in China: an ideological conflict perspective. Children and Youth Services Review 2022;136:106442. [doi: 10.1016/j.childyouth.2022.106442]
40. Jeong Y. Participation, supports, and barriers of Korean children and youth with and without disabilities in the school environment. Disability and Rehabilitation 2020;42(12):1667-1674. [doi: 10.1080/09638288.2019.1567836] [Medline: 30714425]
41. Mulligan H, McCoy L, Wilkinson A, Hawke J, Sullivan J, Kirk E, et al. Disability facilitators' views of their work with children/youth with disability in South Africa. Children and Youth Services Review 2017;82:13-18. [doi: 10.1016/j.childyouth.2017.09.001]
42. Teleman B, Vinblad E, Svedberg P, Nygren JM, Larsson I. Exploring barriers to participation in pediatric rehabilitation: voices of children and young people with disabilities, parents, and professionals. International Journal of Environmental Research and Public Health 2021;18(19):10119 [FREE Full text] [doi: 10.3390/ijerph181910119] [Medline: 34639419]
43. Lindly OJ, Chavez AE, Zuckerman KE. Unmet health services needs among US children with developmental disabilities: associations with family impact and child functioning. Journal of developmental and behavioral pediatrics 2016;37(9):712-723 [FREE Full text] [doi: 10.1097/DBP.0000000000000363] [Medline: 27801721]
44. Xia C, Sun M, Li X, Lu C, Gao X, Lu J, et al. Health-related quality of life and related factors among primary caregivers of children with disabilities in shanghai, China: a cross-sectional study. International Journal of Environmental Research and Public Health 2020;17(24):9299 [FREE Full text] [doi: 10.3390/ijerph17249299] [Medline: 33322674]
45. Piškur B, Beurskens AJHM, Jongmans MJ, Ketelaar M, Smeets RJEM. What do parents need to enhance participation of their school-aged child with a physical disability? a cross-sectional study in the Netherlands. Child: Care, Health and Development 2015;41(1):84-92. [doi: 10.1111/cch.12145] [Medline: 24797584]
46. Houtrow AJ, Carle A, Perrin JM, Stein RE. Children with special health care needs on supplemental security income for disability have more health impacts and needs than other children with special health care needs on medicaid. Academic Pediatrics 2020;20(2):258-266. [doi: 10.1016/j.acap.2019.10.014] [Medline: 31698084]
47. Meehan EM, Reid SM, Williams KJ, Freed GL, Sewell JR, Reddihough DS. Medical service use in children with cerebral palsy: the role of child and family factors characteristics. Journal of Paediatrics and Child Health 2016;52(6):621-627. [doi: 10.1111/jpc.13163] [Medline: 27088437]
48. Umat C, Abdul Wahat NH, Che Ross S, Goh BS. Quality of life of parents and siblings of children with cochlear implants. Journal of Otology 2019;14(1):17-21 [FREE Full text] [doi: 10.1016/j.joto.2018.11.004] [Medline: 30936897]

49. He P, Guo C, Luo Y, Wen X, Salas JI, Chen G, et al. Trends in rehabilitation services use in chinese children and adolescents with intellectual disabilities: 2007-2013. Archives of Physical Medicine and Rehabilitation 2017;98(12):2408-2415. [doi: 10.1016/j.apmr.2017.05.007] [Medline: 28610967]
50. Carter B, Bennett CV, Jones H, Bethel J, Perra O, Wang T, et al. Healthcare use by children and young adults with cerebral palsy. Developmental Medicine & Child Neurology 2021;63(1):75-80 [FREE Full text] [doi: 10.1111/dmcn.14536] [Medline: 32314347]
51. Pérez-Ardanaz B, Morales-Asencio JM, León-Campos Á, Kaknani-Uttumchandani S, López-Leiva I, Garcia-Piñero JM, et al. Quality of life and health services utilization for Spanish children with cerebral palsy. Journal of Pediatric Nursing 2020;53:e121-e128. [doi: 10.1016/j.pedn.2020.03.001] [Medline: 32201113]
52. Matsuzawa A, Shiroki Y. Mothers' experiences of care coordination for children with disabilities: a qualitative study. Children (Basel, Switzerland) 2022;9(6):835 [FREE Full text] [doi: 10.3390/children9060835] [Medline: 35740772]
53. Khusaifan SJ, El Keshky MES. Social support as a protective factor for the well-being of parents of children with autism in Saudi Arabia. Journal of Pediatric Nursing 2021;58:e1-e7. [doi: 10.1016/j.pedn.2020.11.014] [Medline: 33317948]
54. Arabiat DH, Whitehead L, Al Jabery M. The 12-year prevalence and trends of childhood disabilities in Australia: findings from the survey of disability, aging and carers. Child: Care, Health and Development 2018;44(5):697-703. [doi: 10.1111/cch.12588] [Medline: 30022510]
55. Caicedo C. Children with special health care needs: child health and functioning outcomes and health care service use. Journal of Pediatric Health Care 2016;30(6):590-598. [doi: 10.1016/j.pedhc.2015.12.003] [Medline: 26832456]
56. Schickedanz A, Dreyer BP, Halfon N. Childhood poverty: understanding and preventing the adverse impacts of a most-prevalent risk to pediatric health and well-being. Pediatric Clinics of North America 2015;62(5):1111-1135. [doi: 10.1016/j.pcl.2015.05.008] [Medline: 26318943]
57. Racine AD. Child poverty and the health care system. Acad Pediatr 2016;16(3 Suppl):S83-S89. [doi: 10.1016/j.acap.2015.12.002] [Medline: 27044708]
58. Dall'Alba L, Gray M, Williams G, Lowe S. Early intervention in children (0–6 years) with a rare developmental disability: the occupational therapy role. Hong Kong Journal of Occupational Therapy 2014;24(2):72-80. [doi: 10.1016/j.hkjot.2014.12.001]
59. Houtrow A, Murphy N, Council On Children With Disabilities. Prescribing physical, occupational, and speech therapy services for children with disabilities. Pediatrics 2019;143(4):e20190285. [doi: 10.1542/peds.2019-0285] [Medline: 30910917]
60. Ziviani J, Feeney R, Rodger S, Watter P. Systematic review of early intervention programmes for children from birth to nine years who have a physical disability. Australian Occupational Therapy Journal 2010;57(4):210-223. [doi: 10.1111/j.1440-1630.2010.00850.x] [Medline: 20854595]
61. Deal LW, Shiono PH. Medicaid managed care and children: an overview. Future Child 1998;8(2):93-104. [Medline: 9782652]
62. Abbott D, Jessiman P. 'It's afforded us a huge flexibility': the impact of 'disabled children's access to childcare pilots' on families with a disabled child in England. Community, Work & Family 2014;17(4):456-466. [doi: 10.1080/13668803.2014.893227]
63. Raghavendra P, Murchland S, Bentley M, Wake-Dyster W, Lyons T. Parents' and service providers' perceptions of family-centred practice in a community-based, paediatric disability service in Australia. Child: Care, Health and Development 2007;33(5):586-592. [doi: 10.1111/j.1365-2214.2007.00763.x] [Medline: 17725781]
64. Wang KY. The care burden of families with members having intellectual and developmental disorder: a review of the recent literature. Current Opinion in Psychiatry 2012;25(5):348-352. [doi: 10.1097/YCO.0b013e3283564248] [Medline: 22744406]
65. Kasparian NA, Kovacs AH. Quality of life and other patient-reported outcomes across the life span among people with Fontan palliation. The Canadian Journal of Cardiology 2022;38(7):963-976. [doi: 10.1016/j.cjca.2022.04.025] [Medline: 35525399]
66. Nuri RP, Aldersey HM, Ghahari S. Needs of families of children with cerebral palsy in Bangladesh: a qualitative study. Child Care Health Dev 2019 Jan;45(1):36-44. [doi: 10.1111/cch.12624] [Medline: 30304759]
67. Hanga K, DiNitto D, Leppik L. Initial assessment of rehabilitation needs using the WHODAS 2.0 in Estonia. Disabil Rehabil 2016;38(3):260-267. [doi: 10.3109/09638288.2015.1036172] [Medline: 25869288]
68. Guidetti D, Spallazzi M, Baldereschi M, Di Carlo A, Ferro S, Rota E Morelli N, Italy Project-Working Group. Post-stroke rehabilitation in Italy: inconsistencies across regional strategies. Eur J Phys Rehabil Med 2014 Jul;50(3):335-341 [FREE Full text] [Medline: 24525622]


## Abbreviations

**GNI:** gross national income
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**WHO:** World Health Organization

*Edited by T de Azevedo Cardoso; submitted 18.06.23; peer-reviewed by R Kalleson, L Lindstrom; comments to author 19.01.24; revised version received 08.02.24; accepted 29.05.24; published 07.08.24.*

*Please cite as:*
*Xie Y, Wu J, Li Y, Liu H, Peng Y, Zhou P, Sun Y, Kang L, Jiang C, Wu H*
*Requirements for and Barriers to Rehabilitation Services for Children With Disabilities in Middle- and High-Income Countries: Scoping Review*
*Interact J Med Res 2024;13:e50047*
*URL: https://www.i-jmr.org/2024/1/e50047*
*doi:10.2196/50047*
*PMID:39110502*

Review

# Examining the Relationships Between Indoor Environmental Quality Parameters Pertaining to Light, Noise, Temperature, and Humidity and the Behavioral and Psychological Symptoms of People Living With Dementia: Scoping Review

Wan-Tai M Au-Yeung[1], PhD; Lyndsey Miller[1], PhD; Chao-Yi Wu[1,2], PhD; Zachary Beattie[1], PhD; Michael Nunnerley[1,3], BS; Remonda Hanna[1,4], BS; Sarah Gothard[1], BS; Katherine Wild[1], PhD; Jeffrey Kaye[1], PhD

[1]Department of Neurology, Oregon Health & Science University, Portland, OR, United States

[2]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

[3]Department of Psychology, Portland State University, Portland, OR, United States

[4]Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, United States

**Corresponding Author:**
Wan-Tai M Au-Yeung, PhD
Department of Neurology
Oregon Health & Science University
3181 SW Sam Jackson Park Road
Mailcode: UHN13
Portland, OR, 97239
United States
Phone: 1 5034946695
Email: auyeungm@ohsu.edu

## Abstract

**Background:** A common challenge for individuals caring for people with Alzheimer disease and related dementias is managing the behavioral and psychological symptoms of dementia (BPSD). Effective management of BPSD will increase the quality of life of people living with dementia, lessen caregivers' burden, and lower health care cost.

**Objective:** In this review, we seek to (1) examine how indoor environmental quality parameters pertaining to light, noise, temperature, and humidity are associated with BPSD and how controlling these parameters can help manage these symptoms and (2) identify the current state of knowledge in this area, current gaps in the research, and potential future directions.

**Methods:** Searches were conducted in the CINAHL, Embase, MEDLINE, and PsycINFO databases for papers published from January 2007 to February 2024. We searched for studies examining the relationship between indoor environmental quality parameters pertaining to light, noise, temperature, and humidity and BPSD.

**Results:** A total of 3123 papers were identified in the original search in October 2020. After an additional 2 searches and screening, 38 (0.69%) of the 5476 papers were included. Among the included papers, light was the most studied environmental factor (34/38, 89%), while there were fewer studies (from 5/38, 13% to 11/38, 29%) examining the relationships between other environmental factors and BPSD. Of the 38 studies, 8 (21%) examined multiple indoor environmental quality parameters. Subjective data were the only source of environmental assessments in 6 (16%) of the 38 studies. The findings regarding the relationship between agitation and light therapy are conflicted, while the studies that examined the relationship between BPSD and temperature or humidity are all observational. The results suggest that when the environmental factors are deemed overstimulating or understimulating for an individual with dementia, the behavioral symptoms tend to be exacerbated.

**Conclusions:** The findings of this scoping review may inform the design of long-term care units and older adult housing to support aging in place. More research is still needed to better understand the relationship between indoor environmental quality parameters and BPSD, and there is a need for more objective measurements of both the indoor environmental quality parameters and behavioral symptoms. One future direction is to incorporate objective sensing and advanced computational methods in real-time assessments to initiate just-in-time environmental interventions. Better management of BPSD will benefit patients, caregivers, and the health care system.

*(Interact J Med Res 2024;13:e56452)* doi:10.2196/56452

**KEYWORDS**

dementia; behavioral and psychological symptoms of dementia; neuropsychiatric symptoms; physical environment; light; noise; temperature; humidity

# Introduction

## Background

Alzheimer disease and related dementias (ADRD) are a major public health challenge. The number of Americans aged ≥65 years with Alzheimer disease was estimated to be 6.7 million in 2023 [1]. In 2023, total payments for health care, long-term care, and hospice services for people aged ≥65 years with dementia were estimated to be US $345 billion in the United States. The number of people with Alzheimer disease in the United States is projected to grow to 13.8 million by 2060 [1].

A major component of the high psychological and financial costs of ADRD is related to addressing the needs of people living with dementia who have behavioral and psychological symptoms of dementia (BPSD), also called neuropsychiatric symptoms. BPSD encompass the challenging behaviors exhibited by people living with dementia that comprise a variety of symptoms that commonly include, but are not limited to, apathy, anxiety, depression, agitation, delusions, hallucinations, motor disturbances, and sleep changes. These symptoms are associated with faster disease progression [2], increased caregiving burden [3], and earlier placement into long-term care settings [4]. Pharmacological intervention is often not effective and is associated with undesirable side effects [5], with nonpharmacological therapy being a first-line approach to management [6].

The causes of BPSD are multiple, ranging from intrinsic neuropathologic factors to human factors such as caregiver interactions and the social-environmental milieu [7]. There is considerable evidence suggesting that environmental factors can affect the progression of different diseases as well as human behaviors [8,9]. The environment as a risk factor for ADRD has been reviewed, suggesting links between increased risks of ADRD and factors such as poor air quality, environmental toxins, and occupation-related exposures [10,11]. However, what has been less clearly examined is the potential direct relationship between person- or residence-level experienced environmental factors and contemporaneous behavioral changes associated with ADRD. This is important because most older people and people living with dementia spend the majority of their time indoors [12].

There has long been debate regarding the definition of the environment [13]. For the purpose of this review, we are focusing on the ambient environment as defined by Harris et al [14]. Factors pertaining to the ambient environment, such as light, noise, temperature, and humidity, contribute to the visual, acoustic, and thermal comforts of the occupants and may be associated with their BPSD [15]; for example, inadequate light has been suggested to be a risk factor for sundowning syndrome (people with late-stage dementia exhibiting more agitated behaviors in the late afternoon and evening) because of the role of light in vision and its role in circadian rhythm modulation [16,17]. Therefore, designing an indoor environment that would address the needs of people living with dementia is paramount and may reduce their BPSD. A first step to creating an optimal indoor environment for people living with dementia with BPSD is to understand which environmental factors are related to BPSD and how they impact the symptoms.

Previous papers have reviewed the effect of the long-term care environment on physical activity levels [18], the achievement of therapeutic goals such as safety and socialization [19], and neuropsychiatric symptoms [20,21]. However, these reviews have primarily focused on the parameters of the built environment (eg, size and layout of spaces), interior design, and occupancy and staffing ratios, rather than aspects of the environmental quality that impact comfort, such as temperature and humidity, noise levels, and lighting. Comfort is especially important to consider for people living with dementia because BPSD represent an expression of discomfort, stress, and unmet needs when insight and communication become more difficult [22]. Nevertheless, anosognosia and aphasia also prevent the subjective measurement of comfort. Thus, it is necessary to understand the impact of various objective aspects of environmental quality that may improve comfort and, in turn, BPSD.

## Objectives

Accordingly, we conducted this scoping review to (1) examine how indoor environmental quality parameters pertaining to light, noise, temperature, and humidity are associated with BPSD and how controlling these parameters can help manage these symptoms and (2) identify the current state of knowledge on the relationship between indoor environmental quality parameters and BPSD, current gaps in the research, and potential future directions.

# Methods

## Overview

We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting guidelines for this scoping review. We conducted a scoping review because we aimed to (1) provide an overview of the broad literature examining the relationship between indoor environmental quality parameters and BPSD and (2) identify gaps and future directions.

The databases used were CINAHL, Embase, MEDLINE, and PsycINFO. A literature search strategy was crafted for each database that would retrieve records containing a combination of appropriate index terms and text words pertaining to the following concepts: dementia (dementia, Alzheimer disease, etc), environmental conditions (heat, temperature, light, humidity, etc), and BPSD (anxiety, depression, apathy, agitation,

pacing, etc). The complete search strategies can be found in Multimedia Appendices 1-3. The original search was performed on October 7, 2020, and updated on April 21, 2022, and February 22, 2024. The inclusion and exclusion criteria are presented in Textbox 1.

Six reviewers (WMA-Y, CW, MN, LM, RH, and KW) completed the screening process, which consisted of two stages: (1) title and abstract screening and (2) full-text screening. First, the title and abstract of each paper were reviewed by 2 reviewers independently to examine whether the studies met the inclusion criteria. The reviewers then met iteratively to reconcile any differences in their decisions regarding whether a particular paper should be included. If the 2 reviewers could not come to an agreement, a third reviewer would help reach a decision. After title and abstract screening was completed, the included papers at this stage were retrieved for full-text screening, and the process was repeated after reading the papers in their entirety.

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

**Inclusion criteria**

- Written in English
- Peer reviewed
- Studies of older adults with Alzheimer disease and related dementias (ADRD), not including mild cognitive impairment
- Relationships between indoor environmental quality parameters, including light, noise, temperature, and humidity and the behaviors of people living with dementia
- Exposure to environmental conditions leading to behavioral change
- Observational or controlled
- All residential settings
- Studies using measurements of people's behaviors and psychological symptoms as 1 of the outcomes
- Studies that include measurements or descriptions of the indoor environmental quality parameters (light, noise, temperature, and humidity)

**Exclusion criteria**

- Review articles, commentaries, opinions, theses, and letters to editor
- Studies on genetics as risk factors for ADRD
- Studies that examine the effect of decoration, furniture arrangement, and home-like feeling on participants' behaviors
- Studies on the environment as a risk factor for developing ADRD
- Studies on horticultural therapy, music therapy, multisensory stimulation therapy, virtual or augmented reality, or outdoor activities
- Studies published before 2006 (the 15-year cutoff was chosen to include articles that best represent the current state of knowledge)

Six reviewers (WMA-Y, MN, ZB, LM, RH, and SG) extracted data from the included papers after the screening process was completed. A survey built in Qualtrics (Qualtrics International Inc) was created to assist this process. The survey listed questions covering study design, year of publication, the region and residential settings in which the research was conducted, sample size, demographics of the sample, which environmental conditions were assessed and how they were examined, and which BPSD were evaluated and how they were measured. Next, the papers were categorized by the examined environmental factor or factors: (1) light level, (2) noise level, (3) temperature or humidity, and (4) multiple indoor environmental quality parameters. If there were unified themes of BPSD examined in the articles (eg, agitation and mood-related symptoms), the study results of each theme would be synthesized by summarizing the counts of different study designs and providing a narrative summary [23]. All study data were also summarized into tables. In doing so, the current state of knowledge, gaps in research, and future directions were identified.

### Ethical Considerations

Institutional review board approval was not required for this scoping review because the research did not involve human participants.

## Results

### Overview

A total of 3123 articles were identified in the search initiated on October 7, 2020. After excluding papers published before 2006 as well as duplicates, the titles and abstracts of 1887 articles were screened, after which 1707 (90.46%) articles were excluded, and 180 (9.54%) were retrieved for full-text screening. Ultimately, of these 180 articles, 26 (14.4%) were included. The search was updated on April 21, 2022, and on February 22, 2024, and 12 articles were added to the final pool; hence, 38 papers were included in this scoping review (Figure 1). Of the 38 papers, 8 (21%) were randomized controlled trials (RCTs), 14 (37%) were quasi-experiments, 5 (13%) were pretest-posttest studies, 6 (16%) were cross-sectional studies, 4 (11%) were case series, and 1 (3%) was a cohort study (refer to Table 1 for more information).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the search and screening results.

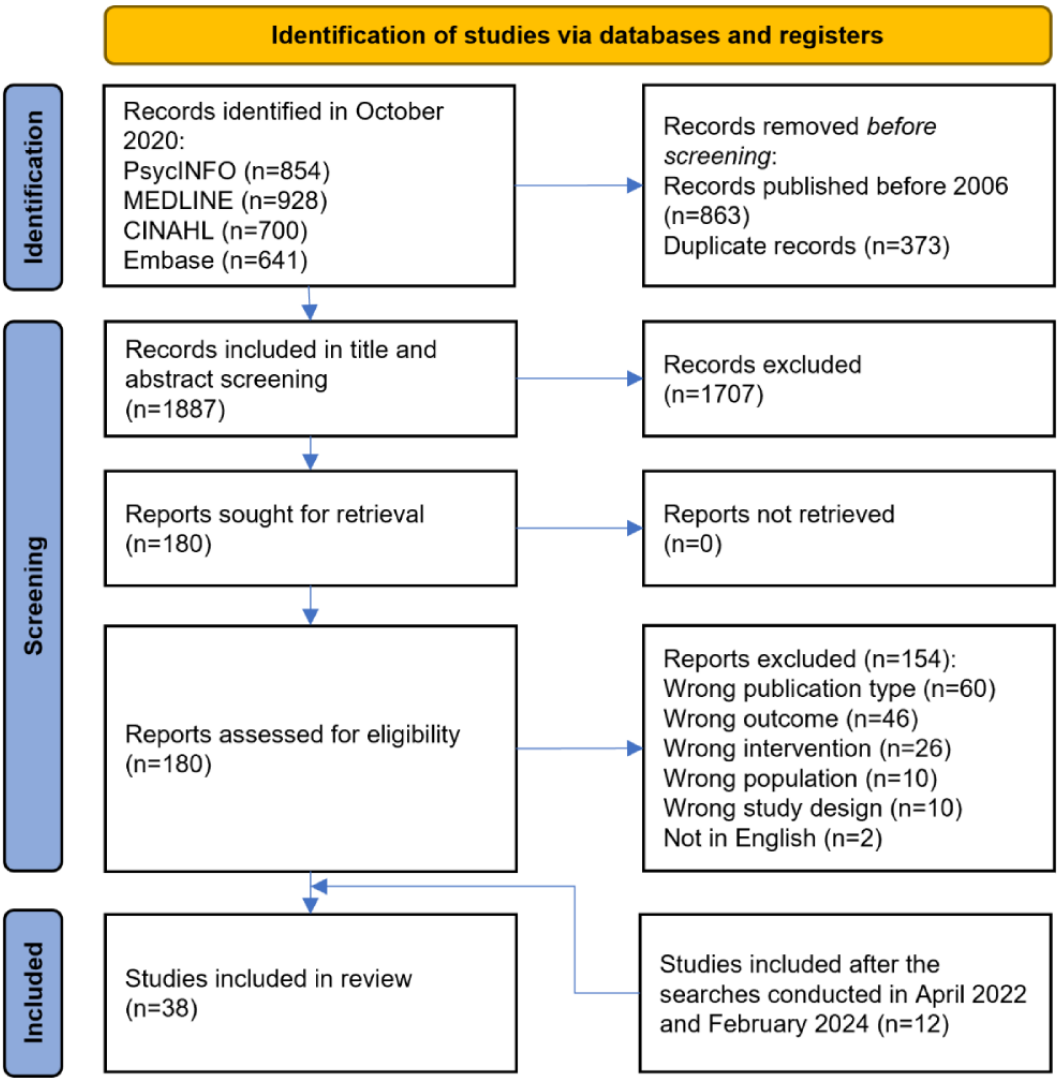


**Table 1.** Summary of residential settings, region, study design, intervention, sample size, participant cognitive status, and participant age for all included studies.

| Study, year | Residential settings | Region | Study design | Intervention (if any) | Sample size | Participant cognitive status | Participant age (y) |
|---|---|---|---|---|---|---|---|
| Algase et al [24], 2010 | Assisted living units, nursing homes | North America | Cross-sectional study | —[a] | M[b]: 28, F[c]: 94 | MMSE[d] score: mean 7.4 (SD 7.2) | Mean 83.7 (SD 6.48; range 68-102) |
| Bankole et al [25], 2020 | Personal homes | North America | Case report or series | — | M: 6, F: 6 (the 12 participants lived with primary caregivers) | All participants had dementia | Mean 79.67 (SD 7.5; range 68-92) |
| Barrick et al [26], 2010 | Memory care units, psychiatric hospital | North America | Quasi-experiment | Four lighting conditions were presented during multiple 3-week intervention periods: bright light in the morning (7-11 Am), bright light in the evening (4-8 PM), bright light all day (7 AM-8 PM), and standard lighting (ie, the baseline condition); bright light was administered at 2000 to 3000 lux, while standard lighting was administered at 500 to 600 lux | M: 35, F: 31 | Cognitive status was measured using the MDS-COGS[e] and MMSE; the distribution of cognitive impairment severity among participants was as follows: mild=3, moderate=18, severe=31, and very severe=14 | North Carolina, United States: <65=6 (13%), 65-79=24 (52%), and ≥80=16 (35%); Oregon, United States: <65=0 (0%), 65-79=4 (20%), and ≥80=16 (80%) |
| Bautrant et al [27], 2019 | Nursing homes | Europe | Pretest-posttest study | Skylike ceiling tiles in part of the shared premises, gradual decrease of the illuminance at night, soothing streaming music, reinforcement of the illuminance during the day, walls painted in light beige, oversized clocks in corridors, and night team clothes color (dark blue) different from that of the day team (sky blue) | M: 2, F: 17 | Severe dementia; MMSE scores were <15 for all patients | ≥65; mean 86.3 (SD 5.4) |
| Bicket et al [28], 2010 | Assisted living units | North America | Cross-sectional study | — | M: 80, F: 246 | Demented: 194, nondemented: 131 | Demented: mean 86.1 (SD 8.1), nondemented: mean 84.3 (SD 9.6) |
| Bromundt et al [29], 2019 | Nursing homes | Europe | Quasi-experiment (balanced crossover, within-participant study) | This was a 17-week trial with a balanced crossover, within- participant study design; the intervention involved exposure to individually timed dawn-dusk simulator over participants' beds for 7-8 weeks | M: 3, F: 17 | S-MMSE[f] score: mean 13.15 (SD 10.30); S-MMSE score at baseline: mean 14.35 (SD 10.51), S-MMSE score at the end of the study: mean 11.95 (SD 11.56) | Mean 85.6 (SD 5.8) |

| Study, year | Residential settings | Region | Study design | Intervention (if any) | Sample size | Participant cognitive status | Participant age (y) |
|---|---|---|---|---|---|---|---|
| Burns et al [30], 2009 | Nursing homes | Europe | Randomized controlled trial | Full-spectrum bright light therapy (10,000 lux) for 2 hours from 10 AM to noon for 2 weeks; the control group was administered standard fluorescent tube light (100 lux) for 2 hours from 10 AM to noon for 2 weeks | M: 16, F: 32 | MMSE score at baseline: 5.1 for the standard lighting therapy group and 6.9 for the bright light therapy group | Standard light therapy group: mean 82.5 (SD 1.5), bright light therapy group: mean 84.5 (SD 1.7) |
| Cohen-Mansfield et al [31], 2010 | Nursing homes | North America | Quasi-experiment (indoor environmental quality parameters were not manipulated) | Presentation of 9 categories of stimuli: live human social, real pet, simulated social, self-identity, inanimate social, music, manipulative, reading, and task- or work-related | M: 42, F: 151 | MMSE score: mean 7.2 (SD 6.3; range 0-23) | Mean 86 (range 60-101) |
| Cohen-Mansfield et al [32], 2011 | Nursing homes | North America | Quasi-experiment (indoor environmental quality parameters were not manipulated) | Presentation of 9 categories of stimuli: live human social, real pet, simulated social, self-identity, inanimate social, music, manipulative, reading and task- or work-related | M: 42, F: 151 | MMSE score: mean 7.2 (SD 6.3; range 0-23) | Mean 86 (range 60-101) |
| Cohen-Mansfield et al [33], 2012 | Nursing homes | North America | Quasi-experiment (indoor environmental quality parameters were not manipulated) | Presentation of 9 categories of stimuli: live human social, real pet, simulated social, self-identity, inanimate social, music, manipulative, reading and task- or work-related | M: 42, F: 151 | MMSE score: mean 7.2 (SD 6.3; range 0-23) | Mean 86 (range 60-101) |
| Cohen-Mansfield [34], 2020 | Nursing homes | North America | Cross-sectional study | — | M: 26, F: 43 | CPS[g] score: mean 4.19 | Mean 86.58 (SD 7.93) |
| Dowling et al [35], 2007 | Nursing homes | North America | Randomized controlled trial | During either morning (9:30-10:30 AM) or afternoon (3:30-4:30 PM), bright light (>2500 lux in gaze direction) was administered Monday through Friday for 10 weeks; Brite Lite IV (Apollo Light Systems, Inc) light boxes were used when necessary to supplement the ambient light | M: 13, F: 57 | MMSE score: mean 7 (SD 7; range 0-23) | Mean 84 (SD 10; range 58-98) |
| Figueiro et al [36], 2014 | Nursing homes | North America | Pretest-posttest study | Low-level *bluish-white* lighting designed to deliver high circadian stimulation during the daytime was installed in 14 nursing home resident rooms for 4 weeks | M: 5, F: 9 | BIMS[h] score: mean 7.7 (SD 2.3) | Mean 86.9 (SD 4.4) |

| Study, year | Residential settings | Region | Study design | Intervention (if any) | Sample size | Participant cognitive status | Participant age (y) |
|---|---|---|---|---|---|---|---|
| Figueiro et al [37], 2019 | Assisted living units, long-term care units | North America | Randomized controlled trial | Crossover design clinical trial administered an all-day active or control lighting intervention for two 4-week periods (separated by a 4-week washout); the active lighting intervention provided high circadian stimulus, and the control intervention provided low circadian stimulus that was below the threshold for activation of the circadian system | M: 16, F: 30 | MMSE score between 4 and 24 or BIMS score between 3 and 12 | Mean 85.1 (SD 7.1) |
| Figueiro et al [38], 2020 | Assisted living units, long-term care units | North America | Pretest-posttest study | Participants were exposed to a single daytime (from approximately 6 AM-8 AM to 6 PM) tailored lighting intervention that provided high levels of circadian-effective light; the light-delivery method (ie, custom-built floor lamps, light boxes, and light tables) was dependent on where individual participants spent most of their day | M: 20, F: 27 | MMSE scores: F=mean 15.2 (SD 3.4) and M=mean 13.6 (SD 2.9); BIMS scores: F=mean 4.0 (SD 1.7) and M=mean 5.6 (SD 2.9) | Mean 85.3 (SD 7.1) |
| Figueiro et al [39], 2023 | Assisted living units, memory care units | North America | Quasi-experiment | With a crossover, placebo-controlled design, 3 different lighting modes (light tables, light trays, and ambient room lighting) were used to deliver high levels of circadian stimulus to the participants' eyes for two 8-week intervention periods in a counter balanced order with a 4-week washout between the study's 2 conditions (dim light control vs active intervention) | M: 3, F: 11 | Moderate to severe ADRD[i] indicated by a GDS[j] score of 4.0 to 6.0; the mean BCRS[k] score was 4.83 (SD 1.01), and the mean GDS score was 4.89 (SD 1.1) | Mean 84.1 (SD 8.9) |
| Garre-Olmo et al [40], 2012 | Nursing homes | Europe | Cross-sectional study | — | M: 37, F 123 | MMSE score: mean 4.1 (SD 6.3); Barthel Index score: mean 10.5 (SD 19.3) | Mean 82.6 (SD 11.6) |

| Study, year | Residential settings | Region | Study design | Intervention (if any) | Sample size | Participant cognitive status | Participant age (y) |
|---|---|---|---|---|---|---|---|
| Hickman et al [41], 2007 | Memory care units, specialized older adult center | North America | Quasi-experiment | Ambient bright light therapy was delivered through a high-intensity, low-glare lighting system installed in the public areas that involved 4 lighting conditions: morning bright light, evening bright light, all-day bright light, and standard lighting. The bright light conditions were delivered at 2000 to 2500 lux, while the standard lighting condition was delivered at 500 to 600 lux | M: 35, F: 31 | Mild to moderate dementia: M=12 (34.3%), F=9 (29%); severe dementia: M=15 (42.9%), F=16 (51.6%); very severe dementia: M=8 (22.9%), F=6 (19.4%) | <65: M=5 (14.3%), F=1 (3.2%); 65-79: M=18 (51.4%), F=10 (32.3%); ≥80: M=12 (34.3%), F=20 (64.5%) |
| Hjetland et al [42], 2021 | Nursing homes | Europe | Randomized controlled trial | A 24-week cluster-randomized, placebo-controlled trial with an intervention involving ambient light administered at 1000 vertical lux and 6000 K from 10 AM to 3 PM, with gradually increasing and decreasing light levels before and after this interval, delivered by Glamox AS | M: 22, F: 47 | MMSE score: mean 6.4 (SD 6.7) | Mean 83.5 (SD 7.1) |
| Jao et al [43], 2015 | Assisted living units, nursing homes | North America | Case report or series | — | M; 9, F: 29 | MMSE score: mean 12.9 (SD 6.5; range 2-23) | Mean 82.7 (SD 6.3; range 68-94) |
| Joosse [44], 2011 | Nursing homes | North America | Cross-sectional study | — | M: 11, F: 42 | MMSE score: mean 6.83 (SD 6.67; range 0 to 19); specifically, 37 (70%) scored <13 (severe dementia), and 16 (30%) scored 13 to 20 (moderate dementia) | Mean 86.53 (SD 9.32; range 61 to 103) |
| Kim et al [45], 2021 | Dementia clinic | Asia | Quasi-experiment | The treatment group (14 participants) sat approximately 60 cm away from a small (136×73×16 mm) light-emitting diode light box for 1 hour each in the morning for 2 weeks, and the control group (11 participants) wore dark, blue-attenuating sunglasses during the 1-hour exposures; the morning light exposure started 9 to 10 hours after each individual's dim light melatonin onset | M: 7, F: 18 | MMSE-KC[l]: treatment group (M: 2, F: 12)=mean 16.36 (SD 5.09), control group (M: 2, F: 12)=mean 16.90 (SD 4.91) | Treatment group: mean 77.36 (SD 5.79), control group: mean 78.55 (SD 7.71) |

| Study, year | Residential settings | Region | Study design | Intervention (if any) | Sample size | Participant cognitive status | Participant age (y) |
|---|---|---|---|---|---|---|---|
| Kolberg et al [46], 2021 | Nursing home dementia units | Europe | Randomized controlled trial | A 24-week cluster-randomized, placebo-controlled trial with an intervention involving ambient light administered at 1000 lux and 6000 K from 10 AM to 3 PM, with gradually increasing and decreasing light levels before and after this interval | M: 22, F: 47 | MMSE score: median 4.0 (IQR 1.0-9.2) | Median 85.0 (IQR 79.0-88.0) |
| Konis et al [47], 2018 | Dementia care communities | North America | Quasi-experiment | The daylight exposure of participants was increased by taking them to the perimeter zone of a daylit room from 8 AM to 10 AM for socialization over a period of 12 weeks; the perimeter zone was the region of the room within 3 m from the windows; it was administered every day throughout the study | M: 21, F: 56 | Alzheimer disease: daylight group=17 (37%) patients, control group=8 (25.8%) patients; frontotemporal dementia: daylight group=1 (2.2%) patient, control group=1 (3.2%) patient; Lewy body dementia: daylight group=0 (0%) patients, control group=1 (3.2%) patient; vascular dementia: daylight group=2 (4.3%) patients, control group=1 (3.2%) patient; dementia not otherwise specified: daylight group=19 (41.3%) patients, control group=17 (54.8%) patients; mild cognitive impairment: daylight group: 6 (13%) patients, control group=2 (6.5%) patients; not specified: daylight group=1 (2.2%) patient, control group=1 (3.2%) patient | Mean 85.3 (SD 7.0) |
| Lee et al [48], 2016 | Long-term care units (Richmond Manor and Maple Manor) | North America | Cohort study | — | M: 6, F: 6 | Early or middle stage of ADRD | Richmond Manor: mean 82.9 (SD 8.9), Maple Manor: mean 77.6 (SD 9.8) |
| Lin et al [49], 2018 | Dementia care centers | Asia | Quasi-experiment | Experimental group participants received 20 minutes of white noise consisting of ocean, rain, wind, and running water sounds between 4 PM and 5 PM daily over a period of 4 weeks, with volume maintained at 55-70 dB; the comparison group received routine care | M: 23, F: 40 | NR[m] | Experimental group: mean 81.57 (SD 10.52), comparison group: mean 79.29 (SD 8.41) |

| Study, year | Residential settings | Region | Study design | Intervention (if any) | Sample size | Participant cognitive status | Participant age (y) |
|---|---|---|---|---|---|---|---|
| Liu et al [50], 2021 | Nursing homes, communities | Asia | Quasi-experiment | Participants in the experimental group were exposed to ambient light at 2500 lux for at least 60 minutes per day from 9 AM to 10 AM, Monday through Friday, over 8 weeks | M: 7, F: 28 | Mild dementia: 10, moderate dementia: 16, severe dementia: 9 | 60-95 |
| Liu et al [51], 2022 | Nursing homes | Asia | Quasi-experiment | Participants in the experimental group were exposed to ambient light at 2500 lux for at least 60 minutes per day from 9 AM to 10 AM, Monday through Friday, over 8 weeks | M: 7, F: 28 | Mild dementia: 10, moderate dementia: 16, severe dementia: 9 | 60-95 experimental group: mean 83.9 (SD 7.1), comparison group: mean 80.2 (SD 7.2) |
| McCurry et al [52], 2011 | Personal homes | North America | Randomized controlled trial | Intervention 1: participants attempted to reach the goal of walking for 30 minutes continuously per day; intervention 2: participants sat in front of a light box (SunRay; The SunBox Company) for 1 hour per day, timed to be within 2-hour window before the participants' habitual bedtime; intervention 3: combination treatment comprising walking, light exposure, and guided sleep education (NITE-AD$^{n}$) | M: 59, F: 73 | MMSE scores: walking group=mean 19.2 (SD 7.7); light group=mean 17.9 (SD 7.0), NITE-AD group=mean 19.1 (SD 5.8), control group=mean 18.7 (SD 6.9) | Walking group: mean 82.2 (SD 8.5), light group: mean 80.6 years (SD 7.3), NITE-AD group: mean 80.0 (SD 8.2), control group: mean 81.2 (SD 8.0) |
| Olsen et al [53], 2016 | Personal homes, nursing homes | Europe | Cross-sectional study | — | M: 67, F: 126 | Nursing home residents (Clinical Dementia Rating Scale): mild dementia=9%, moderate dementia=43.6%, severe dementia=47.4%; home-dwelling people with dementia (Clinical Dementia Rating Scale): mild dementia=43.5%, moderate dementia=47%, severe dementia=5.2% | Nursing home residents: mean 84.6 years (SD 6.50), home-dwelling people with dementia: mean 82.6 (SD 6.84) |
| Onega et al [54], 2016 | Long-term care units | North America | Randomized controlled trial | Individuals were seated approximately 27 inches away from a bright light that delivered 10,000 lux of light (treatment arm) or a low-level light that delivered 250 lux of light (control) for 30 minutes twice a day 5 times a week for 8 weeks | M: 17, F: 43 | MMSE scores ranged from 0 to 22; participants were classified into 3 categories based on the MMSE scores: 7 (11.7%) had mild dementia, 11 (18.3%) had moderate dementia, 42 (70%) had severe dementia | Mean 82.6 (SD 9.6) |

| Study, year | Residential settings | Region | Study design | Intervention (if any) | Sample size | Participant cognitive status | Participant age (y) |
|---|---|---|---|---|---|---|---|
| Riemersma-van der Lek et al [55], 2008 | Group care facilities | Europe | Randomized controlled trial | Random assignment by facility to long-term daily treatment with whole-day bright (approximately 1000 lux) or dim (approximately 300 lux) light and by participant to evening melatonin (2.5 mg) or placebo for a mean of 15 (SD 12) months (maximum period of 3.5 years) | M: 19, F: 170 | Treatment group, placebo: MMSE score=mean 14.3 (SD 7.0); treatment group, light: MMSE score=mean 14.5 (SD 6.2); treatment group, melatonin: MMSE score=mean 15.3 (SD 5.3); treatment group, light+melatonin: MMSE score=mean 14.7 (SD 6.8) | Mean 85.8 (SD 5.5) |
| Saidane et al [56], 2023 | Nursing homes | Europe | Pretest-posttest study | Naturalistic light systems that replicated the spectrum distribution of natural light from dusk to dawn | 5 | NR | NR |
| Sloane et al [57], 2007 | Long-term care units | North America | Quasi-experiment | The study used a cluster-unit crossover experimental design to evaluate 4 ambient lighting conditions: morning (7-11)=high-intensity light, evening (4-8)=high-intensity light, all day (7 AM-8 PM)=high-intensity light, and industry minimum standard lighting | M: 35, F: 31 | Mild to moderate cognitive impairment: 21 (31.8%), severe cognitive impairment: 31 (47%), very severe cognitive impairment: 14 (21.2%) | Mean 79 |
| Son and Kwag [58], 2020 | Center for dementia | Asia | Quasi-experiment | In the experimental group, a walking program with white noise was applied 3 times a week for 4 weeks; white noise was provided by a white noise generator, with volume maintained at 40 to 50 dB; in the control group, only the walking program was applied | M: 16, F: 16 | NR | Range 71-80; experimental group: mean 75.06 (2.82), control group: mean 75.31 (SD 3.07) |
| Tartarini et al [59], 2017 | Nursing homes | Oceania | Case report or series | — | M: 14, F: 7 | Psychogeriatric Assessment Scale scores of ≥10 | Range 61-92 |
| Wahnschaffe et al [60], 2017 | Nursing homes | Europe | Pretest-posttest study | From midwinter on, a ceiling-mounted dynamic lighting system was installed in the common room of a nursing home and programmed to produce high illuminance with higher blue light proportions during the day and lower illuminance without blue light in the evening | M: 5, F: 7 | MMSE score: mean 12.1 (SD 9.2) | Mean 79.1 (SD 11) |

| Study, year | Residential settings | Region | Study design | Intervention (if any) | Sample size | Participant cognitive status | Participant age (y) |
|---|---|---|---|---|---|---|---|
| Wahnschaffe et al [61], 2017 | Nursing homes | Europe | Case report or series | — | M: 1, F: 19 | *ICD-10*[o] criteria: Alzheimer disease=9, vascular dementia=2, frontotemporal dementia=1, Korsakoff syndrome=1, and dementia subtypes (which were not further specified)=7 | Mean 83.8 years (SD 8.8) |

[a]Not applicable.

[b]M: male.

[c]F: female.

[d]MMSE: Mini–Mental State Examination.

[e]MDS-COGS: Minimum Data Set Cognition Scale.

[f]S-MMSE: Standardized Mini–Mental State Examination.

[g]CPS: Cognitive Performance Scale.

[h]BIMS: Brief Interview for Mental Status.

[i]ADRD: Alzheimer disease and related dementias.

[j]GDS: Global Deterioration Scale.

[k]BCRS: Brief Cognitive Rating Scale.

[l]MMSE-KC: Korean version of the MMSE developed as part of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease assessment packet.

[m]NR: not reported.

[n]NITE-AD: Nighttime Insomnia Treatment and Education in Alzheimer's Disease.

[o]ICD-10: International Classification of Diseases, Tenth Revision.

The following sections synthesize the findings of the included studies, categorized by the indoor environmental quality factor or factors examined (Table 2).

**Table 2.** Summary of environmental conditions and behavioral and psychological symptoms of dementia (BPSD) outcomes examined, the subjective and objective methods of assessment, and the main results for all included studies.

| Study, year | Environmental condition or conditions measured or manipulated | Subjective indoor environmental quality parameter assessment (if any) | Objective indoor environmental quality parameter assessment (if any) | Behavioral or psychological outcomes | Subjective BPSD assessment (if any) | Objective BPSD assessment (if any) | Main results |
|---|---|---|---|---|---|---|---|
| Algase et al [24], 2010 | Light level, noise level, temperature, humidity, environmental ambience, crowding | Ambience scale, presence or absence of people within 8 feet of the participant | Gossen Color-Pro 3F light meter (Gossen Color-Pro 3F light meter, Bogen Photo Corp), Quest Technologies Sound Level Meter Model 2400, and RadioShack Thermometer with Indoor Humidity Gauge 63-1013 | Wandering | Observations from videotapes | —[a] | Brighter light, more variation in sound levels, and a higher engaging quality of the environment were associated with wandering, and a higher soothing quality of the environment was associated with periods when wandering did not occur |
| Bankole et al [25], 2020 | Light level, noise level, temperature, humidity, atmospheric pressure | — | Off-the-shelf sensors | Agitation, depression, sleep, quality of life, overall level of BPSD | NPI-Q[b], CMAI-C[c], PSQI[d] | Shimmer3 sensor for people with dementia and Pebble sensor for caregivers in phase 1; Pebble for both people with dementia and caregivers in phase 2 | Temperature, atmospheric pressure, and time showed strong correlations with agitation, which was self-reported by the primary caregivers |
| Barrick et al [26], 2010 | Light level | — | — | Agitation | CMAI[e], observational agitation using the presence or absence of 8 agitated behaviors | — | No therapeutic conditions improved agitation in comparison to standard lighting |
| Bautrant et al [27], 2019 | Light level, skylike ceiling tiles, soothing streaming music, walls painted in light beige, oversized clocks in corridors | — | — | Agitation, wandering, screaming, episodes of BPSD | Number and duration of disruptive BPSD episodes were systematically collected | — | BPSD prevalence reduced after plain environmental rearrangements aimed at improving spatial and temporal orientation were put in place |
| Bicket et al [28], 2010 | Light level | TESS-NH/RC[f] | — | Quality of life, overall level of BPSD, fall risk | NPI[g] | — | The Assisted Living Environmental Quality Score was positively associated with the Alzheimer's Disease–Related Quality of Life score ($P$=.01), while it was strongly negatively associated with NPI total score ($P$<.001) and negatively correlated with fall risk ($P$=.04) |

| Study, year | Environmental condition or conditions measured or manipulated | Subjective indoor environmental quality parameter assessment (if any) | Objective indoor environmental quality parameter assessment (if any) | Behavioral or psychological outcomes | Subjective BPSD assessment (if any) | Objective BPSD assessment (if any) | Main results |
|---|---|---|---|---|---|---|---|
| Bromundt et al [29], 2019 | Light level | — | — | Agitation, mood, circadian rhythm (rest-activity pattern), activities of daily living, quality of life, alertness, verbal interaction, well-being, cheerfulness, memory, disturbing behavior | CMAI, VAS[h] | Wrist actimetry (MotionWatch 8; CamNtech) | Dawn-dusk simulator exposure led to significantly better mood in the morning hours ($P$=.002), while it did not significantly influence circadian parameters and sleep parameters ($P$>.05) |
| Burns et al [30], 2009 | Light level | — | — | Agitation, depression, sleep | CMAI, CSDD[i], MOUSEPAD[j], CRBRS[k], sleep charts | Actiwatch (supplied by CamNtech) | Results suggested that the bright light intervention had a limited effect in reducing agitation, but it improved sleep in older adults with dementia |
| Cohen-Mansfield et al [31] 2010 | Light level, noise level, number of persons in proximity | Environment portion of ABMI[l] | — | Engagement | Observational measurement of engagement | — | Attention and engagement were the best when light level was moderate; engagement was the best when there was a moderate level of noise; attention to the engagement stimulus was significantly higher when there were 4 to 9 people in proximity in comparison to more or fewer people ($P$<.01) |
| Cohen-Mansfield et al [32], 2011 | Light level, noise level, number of persons in proximity | Environment portion of ABMI | — | Agitation | ABMI, CMAI | — | Lighting and background noise did not affect agitation significantly, which may be due to little variation in these variables and their reliance on subjective perception scales |
| Cohen-Mansfield et al [33], 2012 | Light level, noise level, number of persons in proximity | Environment portion of ABMI | — | Pleasure | Lawton Modified Behavior Stream | — | Pleasure was most likely to occur in environments with moderate noise levels |
| Cohen-Mansfield [34], 2020 | Light level, noise level, temperature, location, time of day, total group size | Reported by therapeutic recreation staff | — | Mood, engagement, sleepiness | GOME[l] | — | Background noise and time of day significantly affected outcome variables after controlling for participants' cognitive functioning and group topic; background noise was related with decreased engagement and increased sleepiness; there was little variation concerning temperature and light |
| Dowling et al [35], 2007 | Light level | — | Cal LIGHT 400 calibrated precision light meter for monitoring light levels in gaze direction for each participant | Agitation, depression, overall level of BPSD, appetite or eating disorders, aberrant motor behavior, dysphoria | NPI-NH[n] | — | Bright light therapy was found not to clinically affect neuropsychiatric behaviors |

| Study, year | Environmental condition or conditions measured or manipulated | Subjective indoor environmental quality parameter assessment (if any) | Objective indoor environmental quality parameter assessment (if any) | Behavioral or psychological outcomes | Subjective BPSD assessment (if any) | Objective BPSD assessment (if any) | Main results |
|---|---|---|---|---|---|---|---|
| Figueiro et al [36], 2014 | Light level | — | Daysimeter worn on the wrist to estimate light exposure of each participant | Agitation, depression, circadian rhythm (rest-activity pattern), activities of daily living, sleep | CMAI, CSDD, PSQI | Daysimeter | 300 to 400 lux of a bluish-white light significantly improved sleep efficiency, total sleep time, and global PSQI scores ($P$<.05) and decreased depression (CSDD) and agitation (CMAI) scores ($P$<.05) |
| Figueiro et al [37], 2019 | Light level | — | Daysimeter as a pendant to estimate light exposure of each participant | Agitation, depression, circadian rhythm (rest-activity pattern), sleep, quality of life | CMAI, CSDD, PSQI | Actiwatch 2 (Philips Respironics) | The experimental group had significantly improved PSQI scores compared to the control group ($P$<.001); regarding secondary outcomes, the experimental group had significantly greater improvements in depression and agitation compared to the control group ($P$<.05) |
| Figueiro et al [38], 2020 | Light level | — | Daysimeter as a pendant to estimate light exposure of each participant | Agitation, depression, mood, sleep | CMAI, CSDD, PSQI | Actiwatch 2 (Philips Respironics) | Long-term daily light exposure showed a positive correlation with reduced PSQI scores with increasing efficacy as the study went on; there were significant improvements in depression starting at around week 3; agitation also improved around week 9 |
| Figueiro et al [39], 2023 | Light level | — | — | Depression, circadian rhythm (rest-activity pattern), sleep | CSDD, PSQI, SDI[o] | Actiwatch 2 (Philips Respironics) | Under the active condition, both objective and subjective sleep improved significantly ($P$=.02) |
| Garre-Olmo et al [40], 2012 | Light level, noise level, temperature | — | DT-8820 environment meter (Shenzhen Everbest Machinery Industry Co Ltd) in each participant's bedroom and in the dining room and living room of each nursing home | Affect, mood, pain, quality of life | QUALID[p], NPI-NH, PAIN-AD[q] | — | High temperature in the bedroom was associated with lower quality of life, high noise levels in the living room were associated with low social interactions, and low lighting levels in the bedroom were associated with more severe negative affective mood |
| Hickman et al [41], 2007 | Light level | — | — | Depression | CSDD | — | The treatment had different effects on men and women; study results did not support the use of ambient bright light therapy as a treatment for depressive symptoms in persons with dementia; however, a subpopulation of persons with dementia may benefit from this intervention |
| Hjetland et al [42], 2021 | Light level | — | — | Mood, circadian rhythm (rest-activity pattern), activities of daily living, cognition, sleep, overall level of BPSD | NPI-NH | Actiwatch 2 (Philips Respironics) | There was discrepancy in the subjective and objective sleep measures; actigraphically measured sleep outcomes showed no statistically significant differences between patients in the intervention and control groups; however, better sleep was reported using the SDI in the intervention group |

| Study, year | Environmental condition or conditions measured or manipulated | Subjective indoor environmental quality parameter assessment (if any) | Objective indoor environmental quality parameter assessment (if any) | Behavioral or psychological outcomes | Subjective BPSD assessment (if any) | Objective BPSD assessment (if any) | Main results |
|---|---|---|---|---|---|---|---|
| Jao et al [43], 2015 | Light level, noise level, environmental ambience, crowding, staff familiarity, environmental stimulation | PEAR[r]-Environment subscale | Gossen Color-Pro 3F light meter (Bogen Photo Corp) and the Quest Technologies Sound Level Meter | Apathy | PEAR-Apathy subscale | — | Among the 6 characteristics of environmental stimulation, stimulation clarity and stimulation strength were the only 2 significant factors affecting apathy scores; ambience, crowding, staff familiarity, light, and sounds did not show significant effects on apathy |
| Joosse [44], 2011 | Noise level, number of persons in proximity, spatial environment | — | SoundPro DL 2 (Quest Technologies) sound level meter | Agitation | Wisconsin Agitation Inventory | — | After controlling for potential confounding variables of mental status, hearing impairment, and visual impairment, sound was found to contribute to agitation |
| Kim et al [45], 2021 | Light level | — | — | Depression, sleep, overall level of BPSD | KNPI-Q[s], CSDD-K[t], PSQI, VAS-GV[u], VAS-GA[v] | Actiwatch 2 (Philips Respironics) | The study findings suggest that morning blue-enriched light therapy has a benefit in improving sleep and cognitive function in patients with Alzheimer disease |
| Kolberg et al [46], 2021 | Light level | — | — | Agitation, affect, anxiety, delusion, depression, hallucination, mood, circadian rhythm (rest-activity pattern), sleep, overall level of BPSD | CSDD, NPI-NH | — | Compared to the control group, the intervention group had a larger reduction on the composite scores of both the CSDD and the NPI-NH, as well as on the NPI-NH Affect subsyndrome and the CSDD mood-related signs subscale at follow-up after 16 weeks |
| Konis et al [47], 2018 | Light level | — | Digital charge-coupled device spectrometer | Depression, overall level of BPSD | CSDD, NPI-NH | — | Participants in the daylight intervention experienced an average decrease over the trial in the NPI-NH and CSDD scores, while the control participants showed average but nonsignificant increases in both NPI-NH and CSDD scores; difference in outcome changes of the intervention group achieved statistical significance for the CSDD but not for the NPI-NH |
| Lee et al [48], 2016 | Number of persons in proximity, spatial environment, social environment, crowding, light level | TESS-NH[w] | — | Mood, social engagement | MOSES[x] | — | Residents who stayed in the traditional, large-scale unit showed significantly worse decline in irritable behaviors compared to those who stayed in the small-scale, home-like unit across time; the small-scale, home-like unit had significantly better ratings in stimulation (lighting, visual and tactile stimulation, and noise) as well as personalization according to the TESS-NH |

| Study, year | Environmental condition or conditions measured or manipulated | Subjective indoor environmental quality parameter assessment (if any) | Objective indoor environmental quality parameter assessment (if any) | Behavioral or psychological outcomes | Subjective BPSD assessment (if any) | Objective BPSD assessment (if any) | Main results |
|---|---|---|---|---|---|---|---|
| Lin et al [49], 2018 | Noise level | — | TES-1350A sound level meter | Agitation | CMAI | — | The research results showed that agitated behaviors decreased significantly in the experimental group; in the comparison group, agitated behaviors decreased insignificantly |
| Liu et al [50], 2021 | Light level | — | — | Overall level of BPSD | NPI | Accelerometer (XA-5) | The NPI scores, which were derived using generalized estimating equation with medication (benzodiazepines) as a covariate, were significantly reduced by the fifth and ninth weeks; a lower NPI score indicates less severe BPSD |
| Liu et al [51], 2022 | Light level | — | — | Circadian rhythm (rest-activity pattern), sleep | Sleep charts | Accelerometer (XA-5) | The experimental group showed significantly improved sleep efficiency, sleep duration, and awakening time from baseline to the fifth and ninth week, which was higher than the improvement in the comparison group; the number of nighttime awakenings decreased in the experimental and comparison groups; for circadian rhythm, the experimental group showed significant improvement in sleep onset and sleep offset, which were higher than the improvements in the comparison group |
| McCurry et al [52], 2011 | Light level | — | — | Sleep | SDI | Micro Mini Motionlogger actigraph (Ambulatory Monitoring, Inc) | This study found that exposure to light or increase in walking can have a benefit regarding nighttime sleep; this type of treatment is implementable by caregivers to increase the well-being of those in long-term care facilities; patients who adhered to treatment best saw the best results |
| Olsen et al [53], 2016 | Light level | — | ActiSleep+ (ActiGraph) | Sleep, quality of life, social contact, activity | — | ActiSleep+ (ActiGraph) | People living with dementia in nursing homes had lower quality of life in comparison to those at home, while nursing homes were associated with lower light level; however, in a regression model, residency was the only significant predictor for predicting quality of life in people with moderate dementia after controlling for confounders, including exposure to light level |

| Study, year | Environmental condition or conditions measured or manipulated | Subjective indoor environmental quality parameter assessment (if any) | Objective indoor environmental quality parameter assessment (if any) | Behavioral or psychological outcomes | Subjective BPSD assessment (if any) | Objective BPSD assessment (if any) | Main results |
|---|---|---|---|---|---|---|---|
| Onega et al [54], 2016 | Light level | — | — | Agitation, depression | CMAI, PAS[y], BARS[z], CSDD, DSAOA[aa], DMAS-17[ab] | — | Those who were exposed to bright light over a 2-week period showed a significant improvement in terms of the levels of depression and agitation compared to those who experienced the placebo bright light exposure |
| Riemersma-van der Lek et al [55], 2008 | Light level | — | — | Agitation, affect, depression, mood, MOSES, overall level of BPSD, standardized scales for cognitive and noncognitive symptoms, limitations of activities of daily living, and adverse effects assessed every 6 months | NPI, MOSES, CSDD, PG-CARS[ac], PGCMS[ad], CMAI | Actiwatch (supplied by CamNtech) | Light was shown to help reduce depressive symptoms among the participants by 1.5 points on the CSDD; light in combination with melatonin attenuated aggressive behaviors by 3.9 points on the CMAI and increased sleep efficiency by 3.5% |
| Saidane et al [56], 2023 | Light level | — | — | Agitation | CMAI-inspired score | — | Overall, the frequency of agitation-associated behaviors was reduced by 71.2% after the intervention |
| Sloane et al [57], 2007 | Light level, social environment | — | — | Circadian rhythm (rest-activity pattern), sleep | Sleep charts | — | Those exposed to morning and all-day light showed a significant increase in nighttime sleep duration; circadian rhythm showed significant acrophase shifting; effects on daytime sleepiness were inconsistent |
| Son and Kwag [58], 2020 | Noise level, distracted environment | — | — | Anxiety, fear of falling, walking time | State-Trait Anxiety Inventory–Korean Version | — | White noise during walking was shown to positively decrease the state anxiety and fear of falling in walking among older adults with mild dementia |
| Tartarini et al [59], 2017 | Temperature | — | iButton data loggers (Maxim Integrated) | Agitation | CMAI | — | The CMAI scores increased when residents were exposed to relatively cold or warm indoor temperatures at a statistically significant level; the level of agitation was also significantly correlated to the duration of the increased or decreased temperature |
| Wahnschaffe et al [60], 2017 | Light level | — | Spectroradiometer (Specbos 1201; JETI GmbH) | Agitation, circadian rhythm (rest-activity pattern) | CMAI | Actiwatch (supplied by CamNtech) | The comparison of CMAI sum-scores between assessments before and during the intervention yielded significant differences with decreased agitated behavior after installation of the dynamic lighting system ($P<.05$) but not for rest-activity patterns |

| Study, year | Environmental condition or conditions measured or manipulated | Subjective indoor environmental quality parameter assessment (if any) | Objective indoor environmental quality parameter assessment (if any) | Behavioral or psychological outcomes | Subjective BPSD assessment (if any) | Objective BPSD assessment (if any) | Main results |
|---|---|---|---|---|---|---|---|
| Wahnschaffe et al [61], 2017 | Light level, social environment | — | Local weather data | Circadian rhythm (rest-activity pattern) | — | Actiwatch (supplied by CamNtech) | Nocturnal rest, which was proxied by average activity level during five least active hours, was significantly predicted by cloud amount and day length in the highest number of participants (11 out of 20) among all examined rest-activity-related variables, which included interdaily stability, intradaily variability, and relative amplitude |

[a]Not applicable.
[b]NPI-Q: Neuropsychiatric Inventory Questionnaire.
[c]CMAI-C: Cohen-Mansfield Agitation Inventory–Community form.
[d]PSQI: Pittsburgh Sleep Quality Index.
[e]CMAI: Cohen-Mansfield Agitation Inventory.
[f]TESS-NH/RC: Therapeutic Environment Screening Survey for Nursing Homes and Residential Care.
[g]NPI: Neuropsychiatric Inventory.
[h]VAS: visual analog scale.
[i]CSDD: Cornell Scale for Depression in Dementia.
[j]MOUSEPAD: Manchester and Oxford Universities Scale for the Psychopathological Assessment of Dementia.
[k]CRBRS: Crichton Royal Behavior Rating Scale.
[l]ABMI: Agitation Behavior Mapping Instrument.
[m]GOME: Group Observational Measurement of Engagement.
[n]NPI-NH: Neuropsychiatric Inventory–Nursing Home version.
[o]SDI: Sleep Disorders Inventory.
[p]QUALID: Quality of Life in Late-Stage Dementia.
[q]PAIN-AD: Pain Assessment in Advanced Dementia.
[r]PEAR: Person-Environment Apathy Rating.
[s]KNPI-Q: Korean version of the Neuropsychiatric Inventory Questionnaire.
[t]CSDD-K: Korean version of the Cornell Scale for Depression in Dementia.
[u]VAS-GV: visual analog scale–global vigor.
[v]VAS-GA: visual analog scale–global affect.
[w]TESS-NH: Therapeutic Environment Screening Survey for Nursing Homes.
[x]MOSES: Multidimensional Observational Scale for Elderly Subjects.
[y]PAS: Pittsburgh Agitation Scale.
[z]BARS: Behavioral Activity Rating Scale.
[aa]DSAOA: Depressive Symptom Assessment for Older Adults.
[ab]DMAS-17: Dementia Mood Assessment Scale, 17 items.
[ac]PGCARS: Philadelphia Geriatric Center Affect Rating Scale.
[ad]PGCMS: Philadelphia Geriatric Center Morale Scale.

## Light

### *Overview*

Of the 38 included papers, 34 (89%) examined light as a factor associated with BPSD. Of these 34 studies, 20 (59%) were conducted in North America [24-26,28,31-39,41,43,47,48,52,54,57], 3 (9%) in Asia [45,50,51], and 11 (32%) in Europe [27,29,30,40,42,46,53,55,56,60,61]. Of the 34 studies, 31 (91%) were conducted in long-term care settings, 2 (6%) were conducted in the participants' personal residence [25,52], and 1 (3%) was conducted in both the personal residence and the long-term care setting [53]. The median sample size was 63 (IQR 20-122; range 5-326) participants, and the participants' average age ranged from 73 to 87 years. Light was measured using a variety of light meters or subjective perception scales such as the Therapeutic Environment Screening Survey for Nursing Homes [62] and the environment portion of the Agitation Behavior Mapping Instrument (ABMI) [63].

### *Agitation*

Of the 34 studies, 16 (47%) examined the effect of light on agitation either as a syndrome or related to individual agitated behaviors such as wandering or screaming [24-27,29,30,32,35-38,46,54-56,60]. Of these 16 studies, 14 (88%) examined the effect of changing light exposure in RCTs (n=6, 43%) [30,35,37,46,54,55] or quasi-experimental study designs (n=8, 57%) [26,27,29,32,36,38,56,60], while 2 (12%) examined the effect of light on agitated behaviors in a purely observational or natural history study [24,25]. Among the light therapies studied were bright light therapy [30,35,46,54,55] and outdoor environment simulations, such as dawn-dusk simulation [29,56]. A wide range of light intensities was deployed in these light therapies, ranging from 1000 lux [36] to 10,000 lux [54]. The duration and time of day of application of these light therapies were heterogeneous as well; for example, in the study by Onega et al [54], bright light therapies were applied twice a day (once in the morning and once in the afternoon), while in the study by Kolberg et al [46], light therapy was applied from 10 AM to 3 PM with varying light levels before and after this interval. Due to the heterogeneity of the light therapies and study designs, direct comparison across the studies was infeasible. Among these 16 studies, in addition to the multiple approaches to applying light therapy, there was also heterogeneity in the way agitation was assessed. The Cohen-Mansfield Agitation Inventory [64], which measures the frequencies of 29 different agitated behaviors, was the scale used most commonly to assess the level of agitation in people living with dementia across weeks. Other scales used included the Pittsburgh Agitation Scale [65] and the ABMI [63], both of which are based on direct observations of the patient.

The 6 RCTs, which ranged in study duration from 8 weeks to 15 months, used bright light [30,35,37,46,54,55]. However, the results from the RCTs were conflicting. The symptoms of agitation or restlessness were reduced in 3 (50%) of the 6 RCTs [37,54,55] but not in the remaining 3 (50%) [30,35,46]. The lack of improvement in agitation after bright light exposure in these studies [30,35,46] could be due to measurement errors in the subjective assessment of either light level or level of agitation or because of changes in medications not being tracked, as suggested by the authors.

The additional 10 light studies included 8 (80%) quasi-experimental studies [26,27,29,32,36,38,56,60] and 2 (20%) observational studies [24,25]. Of the 8 quasi-experimental studies, 5 (62%) reported that light therapy could reduce agitated behaviors [27,36,38,56,60], while 3 (38%) showed no effect [26,29,32]. Of these 3 studies, 1 (33%) did not yield significant results in part due to a lack of variation in light level [32]. Of the 2 observational studies, 1 (50%) showed that brighter light was associated with wandering [24].

The study by Figueiro et al [38] showed that the effects of light therapy on patients' agitation varied depending on the participants' cognitive status. The study, which examined the effect of long-term, all-day exposure to circadian-effective light on the sleep, mood, and behavior in persons with dementia, revealed that the lighting intervention was more effective for people with severe dementia than for those with mild or moderate dementia in reducing agitation [38].

### *Disturbed Sleep and Circadian Rhythm*

Of the 17 studies examining the relationship between light and sleep or circadian rhythm [25,29,30,34,36-39,42,45,46,51-53,57,60,61], 5 (29%) were RCTs [30,37,42,46,52], 5 (29%) were quasi-experimental studies [29,39,45,51,57], 3 (18%) were pretest-posttest studies [36,38,60], 2 (12%) were cross-sectional studies [34,53], and 2 (12%) were case series studies [25,61]. All studies that administered light therapy led to better sleep outcomes such as higher sleep efficiency [36], longer nighttime sleep duration [57], fewer awakenings [51], higher circadian rhythm stability [51], or lower Pittsburgh Sleep Quality Index scores indicating better sleep quality [37].

There can be discrepancies between self-reported sleep quality and objectively measured sleep quality. In the study by Hjetland et al [42], ambient bright light treatment was administered to nursing home patients with dementia in a placebo-controlled RCT. While better sleep was reported using the Sleep Disorders Inventory in the bright light therapy group than in the standard lighting group at weeks 16 and 20, actigraphically measured sleep outcomes showed no statistically significant differences between the 2 groups.

### *Mood-Related Symptoms*

Of the 13 studies involving lighting that addressed the relationship between symptoms of depression in people living with dementia and light in the environment [25,30,35-39,41,45-47,54,55], 6 (46%) were RCTs [30,35,37,46,54,55], 4 (31%) were quasi-experimental studies [39,41,45,47], 2 (15%) were pretest-posttest studies [36,38], and 1 (8%) was a case series [25]. A decrease in depressive symptoms was reported in 9 (75%) of the 12 lighting intervention studies [30,36-39,46,47,54,55]; for example, in the RCT described in the study by Onega et al [54], those who were exposed to bright light over a 2-week period showed a significant improvement in depressive symptoms (*P* values ranging from .001 to .017 for DSAOA, DMAS-17, and Cohen-Mansfield Agitation Inventory [CSDD]) while those in the placebo control group showed no improvements in their depressive symptoms. In the quasi-experimental study by Konis et al [47], the participants in a daylight intervention experienced a decrease over the trial in the Cornell Scale for Depression in Dementia (CSDD) score, while the control participants showed a statistically nonsignificant increase in the CSDD score (a higher CSDD score indicates more severe depression). In a pretest-posttest experimental study by Figueiro et al [36], 300 to 400 lux of a blueish-white light was found to significantly decrease depressive symptoms as measured by the CSDD. Some of the studies (2/12, 17%) reported that the efficacy of a lighting intervention is dependent on participants' characteristics. In the study by Hickman et al [41], analysis indicated a sex-by-treatment interaction. Significant sex differences were found in CSDD scores in response to evening light, all-day light, and standard lighting. Male participants experienced significantly more depressive symptoms under morning light than under standard lighting (*P*=.007). By contrast, female participants experienced fewer depressive symptoms under

morning light than under standard lighting, although this result was not statistically significant (*P*=.09).

Similarly, the study by Bromundt et al [29], in which a visual analog scale was used to assess mood, indicated that the improvement in mood from dawn-dusk simulation was dependent on age. In the study, the younger subgroup experienced a stronger positive effect from dawn-dusk simulation on mood [29].

The study by Jao et al [43] examined the effect of light on apathy. The study, which was a case series of people living with dementia in nursing home or assisted living settings, assessed apathy by rating 14 separate 20-minute videos on different days per participant using the Person-Environment Apathy Rating–Apathy subscale [66]. Ambient light (and sound) measurements were collected during the 20-minute video sessions. Light level was found to have had no significant effect on apathy.

The RCT conducted by Kolberg et al [46] in nursing home dementia units with 69 participants examined the effect of light on anxiety and found that the group that received bright light treatment as the intervention had a larger reduction on the composite score of the Neuropsychiatric Inventory–Nursing Home version Affect subsyndrome, which indicated better improvement in depression and anxiety at follow-up after 16 weeks compared to the control group.

## Noise

### *Overview*

Of the 11 studies that examined the relationship between noise and dementia-related behaviors, 5 (45%) were experimental studies [31-33,49,58], 5 (45%) were cross-sectional observational studies [24,25,34,40,44], and 1 (9%) was a case study report [43]. Of these 11 studies, 8 (73%) were conducted in North America [24,25,31-34,43,44], 2 (18%) in Asia [49,58], and 1 (9%) in Europe [40]. Of the 11 studies, 10 (91%) were conducted in long-term care settings, and 1 (9%) was conducted in participants' private home environments [25]. The median sample size was 69 (IQR 38-193; range 10-193) participants, and the participants' average age ranged from 75 to 86 years. Noise was measured using objective sound level meters in 10 (91%) of the 11 studies, while 1 (9%) study [32] used the environment portion of the ABMI [67].

### *Agitation and Wandering*

Of the 5 observational studies, 2 (40%) examined the relationship between noise and agitation among participants with dementia [25,44], and 1 (20%) of the 5 experimental studies aimed to decrease agitation using a white noise intervention [49]. Noise level did not distinguish between episodes of agitation in 1 (50%) [25] of the 2 observational studies, although the study was limited by sample size (10 participants) [25]. In addition, the study by Bankole et al [25] was the only study to examine environmental conditions in the private home setting, which may be distinct from long-term care settings in terms of the effects of noise on agitation. The observational study by Joosse [44], which had 53 participants and used a sound level meter to measure noise levels, found that a higher accumulation of exposure to noise over a day was significantly predictive of agitated behavior in residents with dementia in a long-term care facility [44]. The study by Cohen-Mansfield et al [32] did not find an association between noise and agitation in residents in nursing homes. However, the study was limited by a lack of objective measurement of sound. Finally, in the study by Lin et al [49], an intervention providing daily ambient white noise (55-70 dB) for 20 minutes in the afternoon over a 4-week period significantly decreased the frequency of agitated behavior for participants in the experimental group but not for those in the control group, suggesting that the type of noise, rather than an absence of noise, may be calming.

The observational study by Algase et al [24] examined the effect of noise levels on wandering behavior. In the study of 122 nursing home residents, greater variation in ambient noise level was associated with periods of wandering. There was a significant increase in the risk of wandering (odds ratio 1.09) per SD of variation in ambient noise level [24].

### *Mood-Related Symptoms*

The effect of noise on symptoms related to aspects of mood (apathy, affect, or anxiety) was considered in 7 studies that examined associations with depression [25], apathy [31,43], affect or mood [33,34,40], and anxiety [58]: 3 (43%) quasi-experimental studies [31,33,58], 2 (29%) cross-sectional studies [34,40], and 2 (29%) case series studies [25,43]. When noise levels were moderate, rather than low or high, studies found that participants were more highly engaged (ie, less apathetic) [31,34,40] and exhibited greater pleasure or a more positive attitude [33,34]. High levels of background noise in particular were related to less engagement in 2 (67%) of the 3 studies that examined affect or mood [34,40] but were not related to negative affect, leading the authors to conclude that high background noise may cause participants to become less attuned to their environment and bored or sleepy but not uncomfortable or upset [40]. Another study provided evidence of the need for stimulating sound: Jao et al [43] did not find an association between apathy and overall noise level but did find an association between apathy and a lack of clear, discernible sound stimuli (ie, only chaotic background noise). Finally, signs of anxiety in participants with dementia were decreased with an intervention that applied white noise during a walking program in the experimental group only [58].

## Temperature or Humidity

Of the 5 studies that examined the relationship between temperature or humidity and BPSD [24,25,34,40,59], 3 (60%) were cross-sectional studies [24,34,40], and 2 (40%) were case series [25,59]; furthermore, 3 (60%) were conducted in North America [24,25,34], 1 (20%) in Europe [40], and 1 (20%) in Oceania [59]. Of the 5 studies, 4 (80%) were conducted in nursing homes or assisted living units [24,34,40,59], while 1 (20%) was conducted in participants' private home environments [25]. The median sample size in these studies was 69 (IQR 21-122; range 10-160) participants. The participants' mean age ranged from 79.7 to 86.6 years.

The case series study by Tartarini et al [59] found that the level of agitation of participants with dementia in a nursing home

facility increased significantly when the indoor average temperatures diverged from 22.6 °C. At the same time, the duration of exposure to high temperature (>22.6 °C) and low temperature (<22.6 °C) was linearly correlated with Cohen-Mansfield Agitation Inventory scores. In the cross-sectional studies by Algase et al [24] and Cohen-Mansfield [34], the authors suggested that there was not enough variation in the indoor temperature level or humidity level to identify a relationship with participants' behaviors.

### Multiple Indoor Environmental Quality Parameters

A total of 8 studies [24,25,31-34,40,43] examined multiple indoor environmental quality parameters (refer to Table 2 for the combination of parameters examined in each study). Of these 8 studies, 7 (88%) were conducted in North America [24,25,31-34,43], and 1 (12%) was conducted in Europe [40]; furthermore, 3 (38%) were quasi-experimental studies [31-33], 3 (38%) were cross-sectional studies [24,34,40], and 2 (25%) were case series [25,43]. While the indoor environmental quality parameters—light level and noise level—were measured in the quasi-experimental studies, they were not manipulated. The intervention used in the quasi-experimental studies involved 9 categories of stimuli not related to the indoor environmental quality parameters.

Of the 8 studies, 5 (62%) reported some associations between indoor environmental quality parameters and BPSD. One of the themes that these studies suggested was that both overstimulation and understimulation from indoor environmental quality parameters can worsen BPSD. Algase et al [24] reported that brighter light and more variations in sound levels were associated with wandering. The studies by Cohen-Mansfield et al [31,33] that examined the effect of the physical environment on pleasure, engagement, and mood had similar findings. It was found that in nursing homes, pleasure was most likely to be experienced by participants with dementia in environments with moderate noise levels [33]. In addition, attention and engagement duration in participants with dementia were higher when light was normal in comparison to a dark room, and their attention and attitude were significantly less positive when the lighting in the room was bright than when the lighting was normal [31]. In the same study, all indicators of engagement significantly favored a moderate level of noise over no noise or low noise levels, and engagement duration was significantly longer for a moderate level of noise compared to high and very high levels of noise.

The cross-sectional study by Garre-Olmo et al [40] measured the physical environment in multiple spaces within the nursing home and found that indoor environmental quality parameters (light, temperature, and noise) in individual living spaces in the nursing home were uniquely associated with participants' quality of life, behaviors, and mood as measured with the Quality of Life in Late-Stage Dementia Scale; for example, the authors found that low lighting levels in the bedroom were associated with signs of negative affect, high temperature in the bedroom was associated with lower quality of life experienced by the participants, and high noise levels in the living room were associated with low levels of social interaction.

To examine the relationship between multiple environmental factors and agitation in persons with cognitive decline in their home environments, Bankole et al [25] installed environmental sensors in people's homes that provided continuous data over a 30-day period; in addition, the participants with cognitive decline wore actigraphy devices that provided objective measures of their activity level continuously. Furthermore, 10 caregivers used smartwatches or tablets to report episodes of agitation experienced by the people with cognitive decline in real time, which enabled more direct analysis of the relationship between the behaviors and contemporaneous environmental conditions. Personalized neural networks were built for 6 dyads, which predicted episodes of agitation with light level as a predictor. These networks achieved $F_1$-score values, which are measures of predictive performance in binary classifications, ranging from 86.12% to 97.07%. $F_1$-score values can range from 0% to 100%, with a higher $F_1$-score indicating a better predictive performance. Regarding temperature, the study [25] found that temperature showed strong positive correlation with Teager energy scores from actigraphy device data, which are measures of the aggregate energy of movement [68] and were associated with the agitation exhibited by the people living with dementia.

Of the 8 studies, 3 (38%) failed to identify a link between multiple domains of environmental conditions and BPSD. Jao et al [43] studied the association between apathy in residents with dementia in long-term care facilities and the characteristics of care environments, which included light level, noise level, environmental ambience, crowding, staff familiarity, and environmental stimulation. The authors found that light and sound levels did not show statistically significant effects on apathy [43]. Cohen-Mansfield et al [32] found that lighting and background noise were not associated with levels of agitation in nursing home residents with dementia but that may be due to small variations in these factors. Similarly, Cohen-Mansfield [34] reported no association between noise and mood in people with dementia attending recreational groups.

As seen in Table 2, of the 38 included studies, 6 (16%) relied solely on self-report data for measuring the environment, 13 (34%) used only sensors, and 2 (5%) use both sensors and self-report data, while the environment was not measured but manipulated in 17 (45%) studies. For measurements of BPSD in people living with dementia, of the 38 studies, 36 (95%) used informant-report scales, such as the Neuropsychiatric Inventory Questionnaire; less than half of these studies (n=16, 44%) also used behavioral sensors such as actigraphy to provide objective data.

## Discussion

### Overview

The definition of environment encompasses several major elements: physical environment (eg, light level, noise level, temperature, and humidity), social environment (eg, number of people in proximity), and built environment (eg, furnishings). The focus of this scoping review is on the effects of the indoor environmental quality parameters pertaining to light, noise,

temperature, and humidity on BPSD. Given the fact that most older individuals, including people living with dementia, spend most of their time indoors, better understanding of how these physical indoor environmental quality parameters may affect behavioral health is critical. Overall, this review outlines preliminary evidence on the notable linkages between indoor environmental quality parameters and BPSD. Further research remains to be conducted in the field.

### Principal Findings

The evidence base suggests that light during the day is helpful in modulating circadian rhythm, as well as alleviating sleep disturbance and mood-related disorders. Overall, overstimulation and understimulation of environmental factors can lead to challenging behaviors in people living with dementia; for example, when the environment exhibits high levels or low levels of noise, people living with dementia may exhibit more challenging or disturbing behaviors [44,49]. Similarly, challenging behaviors may be associated with the ambient temperatures experienced by the people living with dementia because there was more agitation observed when the environment was either relatively too hot or too cold for the person living with dementia [59]. Such findings are consistent with the existing models that explain the causes of BPSD, including the unmet needs model [69] and the progressively lowered stress threshold model [70]. Ultimately, finding the proper balance of environmental factors for providing the most personal comfort and quality of life for people living with dementia and caregivers may help to minimize the occurrences of BPSD.

### Future Directions of Research

Light is the most researched indoor environmental quality parameter. Most studies on light therapy have shown that it improves sleep and mood in people living with dementia. However, the findings regarding its effect on agitation are conflicting and warrant further research.

The evidence base generally suggests that when noise levels exceed tolerable ranges, behavioral symptoms become more frequent and severe. Studies indicate that both the time duration of exposure to noise and the type of noise exposure could contribute to behavioral symptoms. Future research should explore these factors further.

Studies examining the effects of temperature or humidity on BPSD are largely observational. There exists a need to better understand how thermal comfort affects BPSD in people living with dementia, especially because there are studies suggesting that older adults may need a higher ambient temperature to attain thermal comfort compared to younger adults [71,72].

Environmental factors were often evaluated using self-reported data, which can be prone to biases. The lack of objective measurement of the environment makes measuring and linking the specific effects of the indoor environmental quality parameters on the behaviors of people living with dementia difficult. Of the 38 included studies, 30 (79%) focused on a single environmental factor, measured over relatively short periods of time, limiting the ability to examine more holistically how the environment affects BPSD. To some degree, this single-domain focus may reflect the complexity as well as the cost of deploying multiple sensor types in multiple indoor spaces. As technology advances, indoor environmental quality parameter sensors are expected to become more affordable and scalable.

Although most environmental exposure occurs indoors, studies need to consider in parallel the outdoor environment in terms of its effects on behavior, which is especially important for studies conducted in personal residences because people also spend time outdoors. Going forward, studies should adopt a more holistic approach by deploying sensors to objectively and longitudinally measure the total environment.

While environmental conditions can be objectively sensed and measured, very often, BPSD of the people living with dementia are either reported by the person affected by dementia or observed and interpreted as occurring by an external observer (usually a caregiver) present when the behavior occurs. In this scoping review, 36 (95%) of the 38 included studies relied on these self-reports or observer reports of BPSD. This type of reporting may contain subjective bias. In addition, a variety of scales have been used for assessing the occurrences of BPSD, which differ in their content, structure, and timing of recall. This likely adds to imprecision in comparing the results across multiple studies. More research is needed to examine how different scales for evaluating BPSD relate to each other and, in turn, to multiple environmental conditions. At the same time, whenever possible, the assessment of BPSD should be corroborated with unobtrusive technology such as actigraphy, which can provide objective data.

As evidenced by multiple studies in this review, the effect of the environment on people living with dementia may be dependent on age, sex, and cognitive state. In addition, manifestations of BPSD may be unique for each person living with dementia as well. Recent work by Iaboni et al [73] showed that the features from wearable multimodal sensors varied in their importance to their predictive models for agitation by both individual and behavior type. Finding the optimal environment for each person living with dementia will help support aging in place, which will improve their quality of life and lower the cost of health care. This may be best achieved by being able to examine the environment and behavior at the individual level with a goal of achieving person-centered assessments of well-being in their environment. The use of continuous multimodal sensing data and frequent self-report facilitates the ability to evaluate and analyze data at the individual or “n-of-1” level [73-75], supporting the principle that each individual is unique. Care providers need to know the unique environments, needs, and preferences of each person with dementia to provide the best care and minimize troubling BPSD. Future research should seek to build effective computational models that will incorporate multiple continuous data streams, both environmental and behavioral, that will predict the occurrences of BPSD so that proactive intervention can be implemented [25].

### Implications for Clinical Practice and Policy

At a macro level, the findings of this review, although they are rudimentary, can help inform the design of long-term care

facilities and residential homes for people living with dementia. Individual assessments of responses to environmental factors can be conducted upon admission; for example, large windows may be used in long-term care facilities to let in sunlight, and this should be supplemented by a dawn-dusk simulation system to address the excessive, or lack of, sunlight in different seasons to modulate residents' sleep and circadian rhythm. In addition, the temperature in long-term care facilities and older adults' homes should be kept at an optimal range that would maximize the comfort of residents. Moreover, quiet flooring and sound-absorbing panels may be used in the long-term care facilities to reduce the noise level; as evidenced by the studies in this review, overly stimulating or noisy environments can increase BPSD. These measures may increase the cost of building and sustaining these long-term care facilities because, for example, maintaining a narrow temperature range can place additional strain on heating, ventilation, and air-conditioning systems in these facilities and may not be energy efficient. However, if the long-term care facilities are more attuned to the needs of people living with dementia, it may result in a decrease in BPSD, improving the quality of life of the people living with dementia and the caregivers as well. This would reduce staff burnout and turnover, ultimately lowering health care costs. Implementing such measures on a broad societal level requires further cost-benefit analysis.

### Limitations

This scoping review has limitations. One of the limitations is that we included studies with heterogeneous designs (eg, observational or experimental), which made the synthesis of results difficult. However, 1 of the aims of this scoping review was to provide an overview of the current state of knowledge regarding the relationship between indoor environmental quality parameters and BPSD. As such, heterogeneous studies were included. Another limitation is that only articles published in English were included. Moreover, only a subset of indoor environmental factors (light level, noise level, temperature, and humidity) were included in this review, while there are other factors relating to BPSD as well, such as psychosocial and biological influences [76]. Notably, this review did not include indoor air quality, another environmental factor that increasingly can be assessed objectively in situ. Air quality has been associated with depression and anxiety in older adults based on spatiotemporal models [77]. A recent study suggested that total volatile organic compound level, among other indoor factors, was significantly correlated with specific areas where behavioral changes occurred in people with moderate to severe dementia in a nursing home [78]. This review only provides evidence for the relationship between a subset of the indoor environmental quality parameters and BPSD, but the relationship likely is best interpreted in a much larger context.

### Conclusions

This review suggests that indoor environmental quality parameters pertaining to light level, noise level, temperature, and humidity may be associated with BPSD. An environment that maximizes the comfort of people living with dementia may decrease their BPSD. Most of the studies (34/38, 89%) in this scoping review pertained to the environmental factor light level, while relatively few studies (5/38, 13%-11/38, 29%) examined the relationship between the remaining indoor environmental quality parameters and BPSD. Among the included studies, there were conflicting findings in the relationship between bright light and agitation, which will need further research. A variety of subjective scales were used to assess the environment and BPSD, which makes synthesizing and comparing results across studies difficult. Going forward, as computational methods and objective sensing technology advance and become more affordable, the behaviors of people living with dementia and their environments should be measured holistically and objectively so that future nonpharmacological intervention can be evidence based.

## Acknowledgments

The authors would like to thank Oregon Health & Science University Library staff for their assistance. The research reported in this publication was supported by the National Institute on Aging of the National Institutes of Health (T32AG055378, K25AG071841, and P30AG066518). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

The Oregon Health & Science University (OHSU), ZB, and JK have a financial interest in Life Analytics, Inc, a company that may have a commercial interest in the results of this research and technology. This potential conflict of interest has been reviewed and managed by OHSU. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Embase search strategy for the original search in 2020.
[DOCX File , 20 KB - ijmr_v13i1e56452_app1.docx ]

Multimedia Appendix 2
CINAHL search strategy for the original search in 2020.
[DOCX File , 13 KB - ijmr_v13i1e56452_app2.docx ]

Multimedia Appendix 3
MEDLINE and PsycINFO search strategy for the original search in 2020.
[DOCX File , 13 KB - ijmr_v13i1e56452_app3.docx ]

## References


1. -. 2023 Alzheimer's disease facts and figures. Alzheimers Dement 2023 Apr;19(4):1598-1695. [doi: 10.1002/alz.13016] [Medline: 36918389]
2. Zahodne LB, Ornstein K, Cosentino S, Devanand D, Stern Y. Longitudinal relationships between Alzheimer disease progression and psychosis, depressed mood, and agitation/aggression. Am J Geriatr Psychiatry 2015 Feb;23(2):130-140 [FREE Full text] [doi: 10.1016/j.jagp.2013.03.014] [Medline: 23871118]
3. Huang SS, Lee M, Liao Y, Wang W, Lai T. Caregiver burden associated with behavioral and psychological symptoms of dementia (BPSD) in Taiwanese elderly. Arch Gerontol Geriatr 2012 Jul;55(1):55-59. [doi: 10.1016/j.archger.2011.04.009] [Medline: 21601931]
4. Balestreri L, Grossberg A, Grossberg GT. Behavioral and psychological symptoms of dementia as a risk factor for nursing home placement. Int Psychogeriatr 2005 Jan 10;12(S1):59-62. [doi: 10.1017/S1041610200006773]
5. Ohno Y, Kunisawa N, Shimizu S. Antipsychotic treatment of behavioral and psychological symptoms of dementia (BPSD): management of extrapyramidal side effects. Front Pharmacol 2019 Sep 17;10:1045 [FREE Full text] [doi: 10.3389/fphar.2019.01045] [Medline: 31607910]
6. Cohen-Mansfield J. Nonpharmacologic interventions for inappropriate behaviors in dementia: a review, summary, and critique. Am J Geriatr Psychiatry 2001;9(4):361-381. [Medline: 11739063]
7. Kales HC, Gitlin LN, Lyketsos CG. Assessment and management of behavioral and psychological symptoms of dementia. BMJ 2015 Mar 02;350(mar02 7):h369 [FREE Full text] [doi: 10.1136/bmj.h369] [Medline: 25731881]
8. Coll CG, Bearer EL, Lerner RM. Nature and Nurture: The Complex Interplay of Genetic and Environmental Influences on Human Behavior and Development. New York, NY: Psychology Press; 2014.
9. Horakova D, Zivadinov R, Weinstock-Guttman B, Havrdova E, Qu J, Tamaño-Blanco M, et al. Environmental factors associated with disease progression after the first demyelinating event: results from the multi-center SET study. PLoS One 2013 Jan 8;8(1):e53996 [FREE Full text] [doi: 10.1371/journal.pone.0053996] [Medline: 23320113]
10. Killin LO, Starr JM, Shiue IJ, Russ TC. Environmental risk factors for dementia: a systematic review. BMC Geriatr 2016 Oct 12;16(1):175 [FREE Full text] [doi: 10.1186/s12877-016-0342-y] [Medline: 27729011]
11. Wang X, Younan D, Millstein J, Petkus AJ, Garcia E, Beavers DP, et al. Association of improved air quality with lower dementia risk in older women. Proc Natl Acad Sci U S A 2022 Jan 11;119(2):e2107833119 [FREE Full text] [doi: 10.1073/pnas.2107833119] [Medline: 34983871]
12. Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, et al. Intelligent systems for assessing aging changes: home-based, unobtrusive, and continuous assessment of aging. J Gerontol B Psychol Sci Soc Sci 2011 Jul 08;66 Suppl 1(Suppl 1):i180-i190 [FREE Full text] [doi: 10.1093/geronb/gbq095] [Medline: 21743050]
13. Gitlin LN, Liebman J, Winter L. Are environmental interventions effective in the management of Alzheimer's disease and related disorders?: A synthesis of the evidence. Alzheimer's Care Today 2003;4(2):85-107 [FREE Full text]
14. Harris PB, McBride G, Ross C, Curtis L. A place to heal: environmental sources of satisfaction among hospital patients. J Applied Social Pyschol 2006 Jul 31;32(6):1276-1299. [doi: 10.1111/j.1559-1816.2002.tb01436.x]
15. Ganesh GA, Sinha SL, Verma TN, Dewangan SK. Investigation of indoor environment quality and factors affecting human comfort: a critical review. Build Environ 2021 Oct;204:108146. [doi: 10.1016/j.buildenv.2021.108146]
16. Guu TW, Aarsland D, Ffytche D. Light, sleep-wake rhythm, and behavioural and psychological symptoms of dementia in care home patients: revisiting the sundowning syndrome. Int J Geriatr Psychiatry 2022 May 26;37(5):5712 [FREE Full text] [doi: 10.1002/gps.5712] [Medline: 35470491]
17. Ramkisoensing A, Meijer JH. Synchronization of biological clock neurons by light and peripheral feedback systems promotes circadian rhythms and health. Front Neurol 2015 Jun 05;6:128 [FREE Full text] [doi: 10.3389/fneur.2015.00128] [Medline: 26097465]
18. Anderiesen H, Scherder EJ, Goossens RH, Sonneveld MH. A systematic review--physical activity in dementia: the influence of the nursing home environment. Appl Ergon 2014 Nov;45(6):1678-1686. [doi: 10.1016/j.apergo.2014.05.011] [Medline: 24968971]
19. Chaudhury H, Cooke HA, Cowie H, Razaghi L. The influence of the physical environment on residents with dementia in long-term care settings: a review of the empirical literature. Gerontologist 2018 Sep 14;58(5):e325-e337. [doi: 10.1093/geront/gnw259] [Medline: 28329827]
20. Zuidema SU, de Jonghe JF, Verhey FR, Koopmans RT. Environmental correlates of neuropsychiatric symptoms in nursing home patients with dementia. Int J Geriatr Psychiatry 2010 Jan 15;25(1):14-22. [doi: 10.1002/gps.2292] [Medline: 19517419]
21. Marquardt G, Bueter K, Motzek T. Impact of the design of the built environment on people with dementia: an evidence-based review. HERD 2014 Oct 01;8(1):127-157. [doi: 10.1177/193758671400800111] [Medline: 25816188]

22. Algase DL, Beck C, Kolanowski A, Whall A, Berent S, Richards K, et al. Need-driven dementia-compromised behavior: an alternative view of disruptive behavior. Am J Alzheimer's Dis 2016 Sep 04;11(6):10-19. [doi: 10.1177/153331759601100603]
23. Pollock D, Peters MD, Khalil H, McInerney P, Alexander L, Tricco AC, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023 Mar 01;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
24. Algase DL, Beattie ER, Antonakos C, Beel-Bates CA, Yao L. Wandering and the physical environment. Am J Alzheimers Dis Other Demen 2010 Jun 08;25(4):340-346 [FREE Full text] [doi: 10.1177/1533317510365342] [Medline: 20378834]
25. Bankole A, Anderson MS, Homdee N, Alam R, Lofton A, Fyffe N, et al. BESI: behavioral and environmental sensing and intervention for dementia caregiver empowerment-phases 1 and 2. Am J Alzheimers Dis Other Demen 2020 Mar 12;35:1533317520906686 [FREE Full text] [doi: 10.1177/1533317520906686] [Medline: 32162529]
26. Barrick AL, Sloane PD, Williams CS, Mitchell CM, Connell BR, Wood W, et al. Impact of ambient bright light on agitation in dementia. Int J Geriatr Psychiatry 2010 Oct 26;25(10):1013-1021 [FREE Full text] [doi: 10.1002/gps.2453] [Medline: 20104513]
27. Bautrant T, Grino M, Peloso C, Schiettecatte F, Planelles M, Oliver C, et al. Impact of environmental modifications to enhance day-night orientation on behavior of nursing home residents with dementia. J Am Med Dir Assoc 2019 Mar;20(3):377-381. [doi: 10.1016/j.jamda.2018.09.015] [Medline: 30455049]
28. Bicket MC, Samus QM, McNabney M, Onyike CU, Mayer LS, Brandt J, et al. The physical environment influences neuropsychiatric symptoms and other outcomes in assisted living residents. Int J Geriatr Psychiatry 2010 Oct 14;25(10):1044-1054 [FREE Full text] [doi: 10.1002/gps.2460] [Medline: 20077498]
29. Bromundt V, Wirz-Justice A, Boutellier M, Winter S, Haberstroh M, Terman M, et al. Effects of a dawn-dusk simulation on circadian rest-activity cycles, sleep, mood and well-being in dementia patients. Exp Gerontol 2019 Sep;124:110641. [doi: 10.1016/j.exger.2019.110641] [Medline: 31252161]
30. Burns A, Allen H, Tomenson B, Duignan D, Byrne J. Bright light therapy for agitation in dementia: a randomized controlled trial. Int Psychogeriatr 2009 Aug 01;21(4):711-721. [doi: 10.1017/S1041610209008886] [Medline: 19323872]
31. Cohen-Mansfield J, Thein K, Dakheel-Ali M, Marx MS. Engaging nursing home residents with dementia in activities: the effects of modeling, presentation order, time of day, and setting characteristics. Aging Ment Health 2010 May 06;14(4):471-480 [FREE Full text] [doi: 10.1080/13607860903586102] [Medline: 20455123]
32. Cohen-Mansfield J, Thein K, Marx MS, Dakheel-Ali M, Murad H, Freedman LS. The relationships of environment and personal characteristics to agitated behaviors in nursing home residents with dementia. J Clin Psychiatry 2012 Mar;73(3):392-399. [doi: 10.4088/JCP.10m06605] [Medline: 22152398]
33. Cohen-Mansfield J, Marx MS, Freedman LS, Murad H, Thein K, Dakheel-Ali M. What affects pleasure in persons with advanced stage dementia? J Psychiatr Res 2012 Mar;46(3):402-406 [FREE Full text] [doi: 10.1016/j.jpsychires.2011.12.003] [Medline: 22208995]
34. Cohen-Mansfield J. The impact of environmental factors on persons with dementia attending recreational groups. Int J Geriatr Psychiatry 2020 Feb 26;35(2):141-146. [doi: 10.1002/gps.5210] [Medline: 31595549]
35. Dowling GA, Graf CL, Hubbard EM, Luxenberg JS. Light treatment for neuropsychiatric behaviors in Alzheimer's disease. West J Nurs Res 2007 Dec 02;29(8):961-975 [FREE Full text] [doi: 10.1177/0193945907303083] [Medline: 17596638]
36. Figueiro MG, Plitnick BA, Lok A, Jones GE, Higgins P, Hornick TR, et al. Tailored lighting intervention improves measures of sleep, depression, and agitation in persons with Alzheimer's disease and related dementia living in long-term care facilities. Clin Interv Aging 2014;9:1527-1537 [FREE Full text] [doi: 10.2147/CIA.S68557] [Medline: 25246779]
37. Figueiro MG, Plitnick B, Roohan C, Sahin L, Kalsher M, Rea MS. Effects of a tailored lighting intervention on sleep quality, rest-activity, mood, and behavior in older adults with Alzheimer disease and related dementias: a randomized clinical trial. J Clin Sleep Med 2019 Dec 15;15(12):1757-1767 [FREE Full text] [doi: 10.5664/jcsm.8078] [Medline: 31855161]
38. Figueiro MG, Sahin L, Kalsher M, Plitnick B, Rea MS. Long-term, all-day exposure to circadian-effective light improves sleep, mood, and behavior in persons with dementia. J Alzheimers Dis Rep 2020 Aug 04;4(1):297-312 [FREE Full text] [doi: 10.3233/ADR-200212] [Medline: 33024938]
39. Figueiro MG, Pedler D, Plitnick B, Zecena E, Leahy S. Tailored lighting intervention (TLI) for improving sleep-wake cycles in older adults living with dementia. Front Physiol 2023 Dec 18;14:1290678 [FREE Full text] [doi: 10.3389/fphys.2023.1290678] [Medline: 38169838]
40. Garre-Olmo J, López-Pousa S, Turon-Estrada A, Juvinyà D, Ballester D, Vilalta-Franch J. Environmental determinants of quality of life in nursing home residents with severe dementia. J Am Geriatr Soc 2012 Jul 15;60(7):1230-1236. [doi: 10.1111/j.1532-5415.2012.04040.x] [Medline: 22702541]
41. Hickman SE, Barrick AL, Williams CS, Zimmerman S, Connell BR, Preisser JS, et al. The effect of ambient bright light therapy on depressive symptoms in persons with dementia. J Am Geriatr Soc 2007 Nov 29;55(11):1817-1824. [doi: 10.1111/j.1532-5415.2007.01428.x] [Medline: 17944896]
42. Hjetland GJ, Kolberg E, Pallesen S, Thun E, Nordhus IH, Bjorvatn B, et al. Ambient bright light treatment improved proxy-rated sleep but not sleep measured by actigraphy in nursing home patients with dementia: a placebo-controlled

randomised trial. BMC Geriatr 2021 May 17;21(1):312 [FREE Full text] [doi: 10.1186/s12877-021-02236-4] [Medline: 34001024]
43. Jao YL, Algase DL, Specht JK, Williams K. The association between characteristics of care environments and apathy in residents with dementia in long-term care facilities. Gerontologist 2015 Jun 08;55 Suppl 1(Suppl 1):S27-S39. [doi: 10.1093/geront/gnu166] [Medline: 26055779]
44. Joosse LL. Do sound levels and space contribute to agitation in nursing home residents with dementia? Res Gerontol Nurs 2012 Jul;5(3):174-184. [doi: 10.3928/19404921-20120605-02] [Medline: 22716652]
45. Kim SJ, Lee SH, Suh IB, Jang J, Jhoo JH, Lee JH. Positive effect of timed blue-enriched white light on sleep and cognition in patients with mild and moderate Alzheimer's disease. Sci Rep 2021 May 13;11(1):10174 [FREE Full text] [doi: 10.1038/s41598-021-89521-9] [Medline: 33986349]
46. Kolberg E, Hjetland GJ, Thun E, Pallesen S, Nordhus IH, Husebo BS, et al. The effects of bright light treatment on affective symptoms in people with dementia: a 24-week cluster randomized controlled trial. BMC Psychiatry 2021 Jul 28;21(1):377 [FREE Full text] [doi: 10.1186/s12888-021-03376-y] [Medline: 34320937]
47. Konis K, Mack WJ, Schneider EL. Pilot study to examine the effects of indoor daylight exposure on depression and other neuropsychiatric symptoms in people living with dementia in long-term care communities. Clin Interv Aging 2018;13:1071-1077 [FREE Full text] [doi: 10.2147/CIA.S165224] [Medline: 29881263]
48. Lee SY, Chaudhury H, Hung L. Effects of physical environment on health and behaviors of residents with dementia in long-term care facilities: a longitudinal study. Res Gerontol Nurs 2016 Mar;9(2):81-91. [doi: 10.3928/19404921-20150709-01] [Medline: 26197068]
49. Lin LW, Weng SC, Wu HS, Tsai LJ, Lin YL, Yeh SH. The effects of white noise on agitated behaviors, mental status, and activities of daily living in older adults with dementia. J Nurs Res 2018 Feb;26(1):2-9. [doi: 10.1097/JNR.0000000000000211] [Medline: 29315202]
50. Liu CR, Liou YM, Jou JH. Pilot study of the effects of bright ambient therapy on dementia symptoms and cognitive function. Front Psychol 2021 Dec 24;12:782160 [FREE Full text] [doi: 10.3389/fpsyg.2021.782160] [Medline: 35002870]
51. Liu CR, Liou YM, Jou JH. Ambient bright lighting in the morning improves sleep disturbances of older adults with dementia. Sleep Med 2022 Jan;89:1-9. [doi: 10.1016/j.sleep.2021.10.011] [Medline: 34844127]
52. McCurry SM, Pike KC, Vitiello MV, Logsdon RG, Larson EB, Teri L. Increasing walking and bright light exposure to improve sleep in community-dwelling persons with Alzheimer's disease: results of a randomized, controlled trial. J Am Geriatr Soc 2011 Aug 28;59(8):1393-1402 [FREE Full text] [doi: 10.1111/j.1532-5415.2011.03519.x] [Medline: 21797835]
53. Olsen C, Pedersen I, Bergland A, Enders-Slegers M, Jøranson N, Calogiuri G, et al. Differences in quality of life in home-dwelling persons and nursing home residents with dementia - a cross-sectional study. BMC Geriatr 2016 Jul 11;16(1):137 [FREE Full text] [doi: 10.1186/s12877-016-0312-4] [Medline: 27400744]
54. Onega LL, Pierce TW, Epperly L. Effect of bright light exposure on depression and agitation in older adults with dementia. Issues Ment Health Nurs 2016 Sep 02;37(9):660-667. [doi: 10.1080/01612840.2016.1183736] [Medline: 27254531]
55. Riemersma-van der Lek RF, Swaab DF, Twisk J, Hol EM, Hoogendijk WJ, van Someren EJ. Effect of bright light and melatonin on cognitive and noncognitive function in elderly residents of group care facilities: a randomized controlled trial. JAMA 2008 Jun 11;299(22):2642-2655. [doi: 10.1001/jama.299.22.2642] [Medline: 18544724]
56. Saidane HA, Rasmussen T, Andersen K, Iversen HK, West AS. An explorative investigation of the effect of naturalistic light on agitation-associated behavior in nursing home residents with dementia: a pilot study. HERD 2023 Apr 26;16(2):146-154. [doi: 10.1177/19375867221146154] [Medline: 36573255]
57. Sloane PD, Williams CS, Mitchell CM, Preisser JS, Wood W, Barrick AL, et al. High-intensity environmental light in dementia: effect on sleep and activity. J Am Geriatr Soc 2007 Oct;55(10):1524-1533. [doi: 10.1111/j.1532-5415.2007.01358.x] [Medline: 17714459]
58. Son SM, Kwag SW. Effects of white noise in walking on walking time, state anxiety, and fear of falling among the elderly with mild dementia. Brain Behav 2020 Dec 09;10(12):e01874 [FREE Full text] [doi: 10.1002/brb3.1874] [Medline: 33034429]
59. Tartarini F, Cooper P, Fleming R, Batterham M. Indoor air temperature and agitation of nursing home residents with dementia. Am J Alzheimers Dis Other Demen 2017 Aug 21;32(5):272-281 [FREE Full text] [doi: 10.1177/1533317517704898] [Medline: 28429641]
60. Wahnschaffe A, Nowozin C, Haedel S, Rath A, Appelhof S, Münch M, et al. Implementation of dynamic lighting in a nursing home: impact on agitation but not on rest-activity patterns. Curr Alzheimer Res 2017 Sep 08;14(10):1076-1083. [doi: 10.2174/1567205014666170608092411] [Medline: 28595522]
61. Wahnschaffe A, Nowozin C, Rath A, Floessner T, Appelhoff S, Münch M, et al. Night-time activity forecast by season and weather in a longitudinal design - natural light effects on three years' rest-activity cycles in nursing home residents with dementia. Int Psychogeriatr 2017 Dec 10;29(12):2071-2080. [doi: 10.1017/S1041610217001235] [Medline: 28689500]
62. Sloane PD, Mitchell CM, Weisman G, Zimmerman S, Foley KM, Lynn M, et al. The therapeutic environment screening survey for nursing homes (TESS-NH): an observational instrument for assessing the physical environment of institutional settings for persons with dementia. J Gerontol B Psychol Sci Soc Sci 2002 Mar 01;57(2):S69-S78. [doi: 10.1093/geronb/57.2.s69] [Medline: 11867668]

63. Cohen-Mansfield J. Conceptualization of agitation: results based on the Cohen-Mansfield Agitation Inventory and the Agitation Behavior Mapping Instrument. Int Psychogeriatr 1996 Jan 07;8 Suppl 3(S3):309-351. [doi: 10.1017/s1041610297003530] [Medline: 9154580]
64. Koss E, Weiner M, Ernesto C, Cohen-Mansfield J, Ferris SH, Grundman M, et al. Assessing patterns of agitation in Alzheimer's disease patients with the Cohen-Mansfield agitation inventory. The Alzheimer's disease cooperative study. Alzheimer Dis Assoc Disord 1997;11 Suppl 2:S45-S50. [doi: 10.1097/00002093-199700112-00007] [Medline: 9236952]
65. Rosen J, Burgio L, Kollar M, Cain M, Allison M, Fogleman M, et al. The Pittsburgh Agitation Scale: a user-friendly instrument for rating agitation in dementia patients. Am J Geriatr Psychiatry 1994 Dec;2(1):52-59. [doi: 10.1097/00019442-199400210-00008] [Medline: 28531073]
66. Jao YL, Algase DL, Specht JK, Williams K. Developing the person-environment apathy rating for persons with dementia. Aging Ment Health 2016 Aug 18;20(8):861-870. [doi: 10.1080/13607863.2015.1043618] [Medline: 25984756]
67. Cohen-Mansfield J, Werner P, Marx MS. An observational study of agitation in agitated nursing home residents. Int Psychogeriatr 1989 Jan 07;1(2):153-165. [doi: 10.1017/s1041610289000165] [Medline: 2491142]
68. Bankole A, Anderson M, Knight A, Oh K, Smith-Jackson TL, Hanson MA, et al. Continuous, non-invasive assessment of agitation in dementia using inertial body sensors. In: Proceedings of the 2nd Conference on Wireless Health. 2011 Presented at: WH '11; October 10-13, 2011; San Diego, CA p. 1-9 URL: https://dl.acm.org/doi/10.1145/2077546.2077548
69. Cohen-Mansfield J, Dakheel-Ali M, Marx MS, Thein K, Regier NG. Which unmet needs contribute to behavior problems in persons with advanced dementia? Psychiatry Res 2015 Jul 30;228(1):59-64 [FREE Full text] [doi: 10.1016/j.psychres.2015.03.043] [Medline: 25933478]
70. Richards KC, Beck CK. Progressively lowered stress threshold model: understanding behavioral symptoms of dementia. J Am Geriatr Soc 2004 Oct 23;52(10):1774-1775. [doi: 10.1111/j.1532-5415.2004.52477.x] [Medline: 15450062]
71. Degroot DW, Kenney WL. Impaired defense of core temperature in aged humans during mild cold stress. Am J Physiol Regul Integr Comp Physiol 2007 Jan;292(1):R103-R108 [FREE Full text] [doi: 10.1152/ajpregu.00074.2006] [Medline: 17197640]
72. van Hoof J, Hensen JL. Thermal comfort and older adults. Gerontechnology 2006;4(4):223-228. [doi: 10.4017/gt.2006.04.04.006.00]
73. Iaboni A, Spasojevic S, Newman K, Schindel Martin L, Wang A, Ye B, et al. Wearable multimodal sensors for the detection of behavioral and psychological symptoms of dementia using personalized machine learning models. Alzheimers Dement (Amst) 2022 Apr 27;14(1):e12305 [FREE Full text] [doi: 10.1002/dad2.12305] [Medline: 35496371]
74. Au-Yeung WT, Miller L, Beattie Z, Dodge HH, Reynolds C, Vahia I, et al. Sensing a problem: proof of concept for characterizing and predicting agitation. Alzheimers Dement (N Y) 2020 Aug 24;6(1):e12079 [FREE Full text] [doi: 10.1002/trc2.12079] [Medline: 32864417]
75. Davidoff H, van den Bulcke L, Vandenbulcke M, de Vos M, van den Stock J, van Helleputte N, et al. Toward quantification of agitation in people with dementia using multimodal sensing. Innov Aging 2022;6(7):igac064 [FREE Full text] [doi: 10.1093/geroni/igac064] [Medline: 36600807]
76. Tible OP, Riese F, Savaskan E, von Gunten A. Best practice in the management of behavioural and psychological symptoms of dementia. Ther Adv Neurol Disord 2017 Aug 19;10(8):297-309 [FREE Full text] [doi: 10.1177/1756285617712979] [Medline: 28781611]
77. Pun VC, Manjourides J, Suh H. Association of ambient air pollution with depressive and anxiety symptoms in older adults: results from the NSHAP study. Environ Health Perspect 2017 Mar;125(3):342-348. [doi: 10.1289/ehp494]
78. Ma C, Guerra-Santin O, Mohammadi M. Exploring the influence of indoor environment and spatial layout on changed behaviours of people with dementia in a nursing home. Build Environ 2024 May;256:111452. [doi: 10.1016/j.buildenv.2024.111452]


## Abbreviations

**ABMI:** Agitation Behavior Mapping Instrument
**ADRD:** Alzheimer disease and related dementias
**BPSD:** behavioral and psychological symptoms of dementia
**CSDD:** Cornell Scale for Depression in Dementia
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
**RCT:** randomized controlled trial

*Edited by G Eysenbach, T de Azevedo Cardoso; submitted 16.01.24; peer-reviewed by P Loveland, SQ Yoong; comments to author 10.02.24; revised version received 27.05.24; accepted 28.06.24; published 09.08.24.*

*Please cite as:*
*Au-Yeung WTM, Miller L, Wu CY, Beattie Z, Nunnerley M, Hanna R, Gothard S, Wild K, Kaye J*
*Examining the Relationships Between Indoor Environmental Quality Parameters Pertaining to Light, Noise, Temperature, and Humidity and the Behavioral and Psychological Symptoms of People Living With Dementia: Scoping Review*
*Interact J Med Res 2024;13:e56452*
*URL: https://www.i-jmr.org/2024/1/e56452*
*doi:10.2196/56452*
*PMID:*

Review

# Debate and Dilemmas Regarding Generative AI in Mental Health Care: Scoping Review

Xuechang Xian[1,2*], PhD; Angela Chang[1,3*], Prof Dr; Yu-Tao Xiang[4], Prof Dr; Matthew Tingchi Liu[5*], Prof Dr

[1]Department of Communication, Faculty of Social Sciences, University of Macau, Macau SAR, China

[2]Department of Publicity, Zhaoqing University, Zhaoqing City, China

[3]Institute of Communication and Health, Lugano University, Lugano, Switzerland

[4]Department of Public Health and Medicinal Administration, Faculty of Health Sciences, University of Macau, Macau SAR, China

[5]Faculty of Business Administration, University of Macau, Macau SAR, China

[*]these authors contributed equally

**Corresponding Author:**
Angela Chang, Prof Dr
Department of Communication
Faculty of Social Sciences
University of Macau
University Avenue
Taipa
Macau SAR, 999078
China
Phone: 86 88228991
Email: wychang@um.edu.mo

## Abstract

**Background:** Mental disorders have ranked among the top 10 prevalent causes of burden on a global scale. Generative artificial intelligence (GAI) has emerged as a promising and innovative technological advancement that has significant potential in the field of mental health care. Nevertheless, there is a scarcity of research dedicated to examining and understanding the application landscape of GAI within this domain.

**Objective:** This review aims to inform the current state of GAI knowledge and identify its key uses in the mental health domain by consolidating relevant literature.

**Methods:** Records were searched within 8 reputable sources including Web of Science, PubMed, IEEE Xplore, medRxiv, bioRxiv, Google Scholar, CNKI and Wanfang databases between 2013 and 2023. Our focus was on original, empirical research with either English or Chinese publications that use GAI technologies to benefit mental health. For an exhaustive search, we also checked the studies cited by relevant literature. Two reviewers were responsible for the data selection process, and all the extracted data were synthesized and summarized for brief and in-depth analyses depending on the GAI approaches used (traditional retrieval and rule-based techniques vs advanced GAI techniques).

**Results:** In this review of 144 articles, 44 (30.6%) met the inclusion criteria for detailed analysis. Six key uses of advanced GAI emerged: mental disorder detection, counseling support, therapeutic application, clinical training, clinical decision-making support, and goal-driven optimization. Advanced GAI systems have been mainly focused on therapeutic applications (n=19, 43%) and counseling support (n=13, 30%), with clinical training being the least common. Most studies (n=28, 64%) focused broadly on mental health, while specific conditions such as anxiety (n=1, 2%), bipolar disorder (n=2, 5%), eating disorders (n=1, 2%), posttraumatic stress disorder (n=2, 5%), and schizophrenia (n=1, 2%) received limited attention. Despite prevalent use, the efficacy of ChatGPT in the detection of mental disorders remains insufficient. In addition, 100 articles on traditional GAI approaches were found, indicating diverse areas where advanced GAI could enhance mental health care.

**Conclusions:** This study provides a comprehensive overview of the use of GAI in mental health care, which serves as a valuable guide for future research, practical applications, and policy development in this domain. While GAI demonstrates promise in augmenting mental health care services, its inherent limitations emphasize its role as a supplementary tool rather than a replacement for trained mental health providers. A conscientious and ethical integration of GAI techniques is necessary, ensuring a balanced approach that maximizes benefits while mitigating potential challenges in mental health care practices.

*(Interact J Med Res 2024;13:e53672)* doi:10.2196/53672

**KEYWORDS**

generative artificial intelligence; GAI; ChatGPT; mental health; scoping review; artificial intelligence; depression; anxiety; generative adversarial network; GAN; variational autoencoder; VAE

# Introduction

## Background

Mental disease is among the top 10 leading causes of global burden [1]. One out of every 2 people worldwide will develop at least 1 mental disease in their lifetime [2]. Depression and anxiety disorders are among the most prevalent mental health conditions, significantly impacting individuals' quality of life. It is estimated that approximately 280 million people worldwide are living with depression, with another 301 million experiencing anxiety [3]. These mental diseases can lead to debilitating symptoms, impairing social functioning, and affecting overall well-being [4,5]. People with depression or anxiety are also at a high risk of suicide. More than 700,000 people take their lives every year, with >77% of global suicides occurring in low- and middle-income countries [6].

Despite concerted efforts to address mental health problems, several challenges persist. Limited access to mental health services, especially in resource-limited countries, resulted in a huge treatment gap in mental health care [7]. The lack of sufficient trained mental health professionals further exacerbates this problem, leading to long waiting times for consultations and inadequate support [8]. In addition, stigma and discrimination surrounding mental health continue to hinder individuals from seeking the help they need. Many individuals feel ashamed or worried about the potential consequences of disclosing their mental health conditions, thus impeding early intervention and treatment [7].

Artificial intelligence (AI) appears as a viable alternative for facilitating accessibility, affordability, and anonymity in psychiatric diagnosis and treatment due to its capability to mimic human cognitive functions to learn and make decisions [9]. This is particularly true in machine learning (ML), which constitutes a crucial foundation for AI and focuses on the development of algorithms to learn patterns from training data. ML models can be broadly classified into 2 types: discriminative and generative [10]. In AI systems, discriminative AI models refer to a subset of AI techniques and models that focus on learning the mapping from input data to output labels or categories [10,11]. This approach aims to classify or predict outcomes based on input features without necessarily understanding the underlying relationships or causality in the data. Examples of discriminative AI include image recognition systems and diagnostic systems [11,12].

In prior research, 5 key domains have been identified using discriminative models for bipolar disorder, which include diagnosis, prognosis, treatment, data-driven research, and clinical support [13]. While discriminative AI has been a crucial component of the AI landscape over the past decade and has reached a relatively advanced stage of development, the field of generative AI (GAI), by contrast, remains in its nascent stage. GAI is a type of ML technology that possesses the ability to automatically generate fresh output data by using the provided input data [10]. In fact, the concept of GAI is not a recent innovation but rather a technology that can be traced back to as early as 1966 with ELIZA, a chatbot designed to simulate conversation with a therapist, serving as an early example [14]. However, it is only in recent years that the culmination of extensive research and developments in AI and ML has resulted in the emergence of advanced GAI systems. Unlike traditional GAI, which primarily relies on pattern-matching or rule-based techniques [15], advanced GAI has shown superior performance in autonomously producing synthetic data, text, images, and even videos that resemble real-world examples [10]. Although such an emerging field presents a novel and exciting technological advancement, limited research has attended to inform the application landscape of GAI in mental health care. This indicates an urgent need to expand our knowledge about the state-of-the-art technologies and illuminate their possibilities in clinical and public mental health improvement.

## Objective

The objective of this scoping review is to synthesize available literature describing the use of GAI for mental health to inform the current state of knowledge in this area. The findings of this review can be used to inform various stakeholders, including researchers, clinicians, and support seekers, about the potential uses and implications of GAI in the field of mental health. To achieve this purpose, the review is divided into 2 sections aiming to provide a more thorough overview of the GAI's implementation in mental health care. The first section offers a brief review of research using traditional GAI approaches. These approaches include rule-based or retrieval-based systems that rely on predefined rules or logic statements to perform tasks or generate responses. This contrasts with the discriminative AI approach that involves training models on labeled data to classify inputs into different categories or predict outcomes. The second section presents an in-depth analysis of studies that leveraged advanced GAI. The in-depth review is responsible for the identification of key use cases for advanced GAI alongside its benefits and challenges, while the brief review informs additional areas in which advanced GAI could be leveraged to facilitate the development of evidence-based interventions that can improve mental health outcomes.

# Methods

## Overview

A scoping review, according to Tricco et al [16], is a methodical strategy for "charting" or "mapping" a larger subject than a systematic review usually tackles. This approach becomes essential when addressing the broad problems posed by patients, physicians, politicians, and other decision makers. Therefore, to achieve the objective of this study, we conducted a scoping

review following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [16].

### Search Strategy

The search for relevant studies was conducted between January 1, 2013, and July 28, 2023. The chosen time range is critical for including the latest findings and approaches, ensuring that the research stays relevant and applicable in the evolving field of GAI development [17]. This period also witnessed the emergence of sophisticated GAI tools, which fundamentally transformed approaches to mental health care, shifting from text-based interactions to more engaging forms of mental health support [17,18]. To cover both general and health care–specific sources, we searched records within 3 reputable international databases (ie, Web of Science, PubMed, and IEEE Xplore). In addition, we expanded our search by including 2 Chinese databases, CNKI, and Wanfang, to broaden the scope of our investigation. By including these diverse data sources, we aimed to capture a wide range of literature on GAI from both global and Chinese perspectives.

Given the novelty of the advanced technology, 2 preprint servers (eg, medRxiv and bioRxiv) were consulted from January to July 2023 to inform the scope of literature delineating the use of GAI models for addressing mental health issues. This is a common practice often used by many researchers to locate other emerging applications not yet captured by peer-reviewed literature [19-22]. In order to include additional studies related to our topic, we also performed a structured search on Google Scholar for the identification of uncovered records, including gray literature, and checked the reference lists of the included studies. The search query mainly consists of 2 components: GAI and relevant terms (eg, *GAI* OR *generative model* OR *ChatGPT*) and mental health and related terms (eg, *mental health* OR *depression* OR *anxiety* OR *bipolar disorder*). The search strategies used varied depending on the characteristics of the selected databases for the search inquiry, which are provided in Multimedia Appendix 1.

This review considers only original research for inclusion. Studies were included if they were empirical, leveraged AI-based technologies to generate new outputs for mental health enhancement, and were published in English or Chinese language. However, articles presented in the form of opinions or reports or not relevant to the production of new content were excluded to ensure the relevance and accuracy of the review.

This scoping review categorizes the GAI into advanced GAI and traditional GAI approaches. Advanced GAI features state-of-the-art AI technology for creating new content, images, audio, video, or codes with tools such as ChatGPT, MidJourney, New Bing, and Dall-E2. These models often require huge computing power including a significant amount of memory, while traditional models mainly rely on predefined rules or retrieval-based algorithms with patterns created by developers on the basis of anticipated user queries [15]. Traditional GAI approaches often have limited flexibility and can only provide responses that have been explicitly programmed into their system [23]. This differentiation was deemed essential due to the extensive volume of literature that was identified during the scoping review process. More details of the categorization of traditional and advanced GAI modeling approaches are listed in Multimedia Appendix 2.

### Data Screening and Extraction

All collected records were first imported into EndNote Software (version 20; Clarivate), a literature management software for data screening and storage. Duplicate records were removed, and the remaining articles were screened for relevance based on the information provided in titles and abstracts by one reviewer. This was followed by the other reviewer performing the second stage of screening, which evaluates the full text of articles based on the inclusion and exclusion criteria. Regular scrutiny was held during the screening process to deliberate and address any ambiguities or disagreements and achieve a consensus among both reviewers [24]. A continual reassessment of the understanding of the screening criteria was undertaken. In instances where questions arise, efforts were made to retrace our steps and ascertain the accurate and consistent application of the criteria to guarantee that the screening process maintained a uniform and unbiased standard.

After that, pertinent data were extracted, and a thorough analysis of the application of advanced GAI models for mental health care was carried out. The first author methodically gathered these data, which contain details about the mental health issues addressed, the targeted population, the application setting, the use case, the results, the study phase, the data source, the type of approach, the delivery mechanism, and so on. The extracted data were then synthesized through a narrative method. Our goal was to categorize the GAI studies based on the extracted data. For this purpose, we modified various taxonomies found in existing literature [18,25]. We compiled the characteristics of the studies into a table and provided a narrative description. Following this, we presented an overview of the features in the studies included. The PRISMA checklist was provided in Multimedia Appendix 3 to show the completeness and transparency of the review's reporting [16].

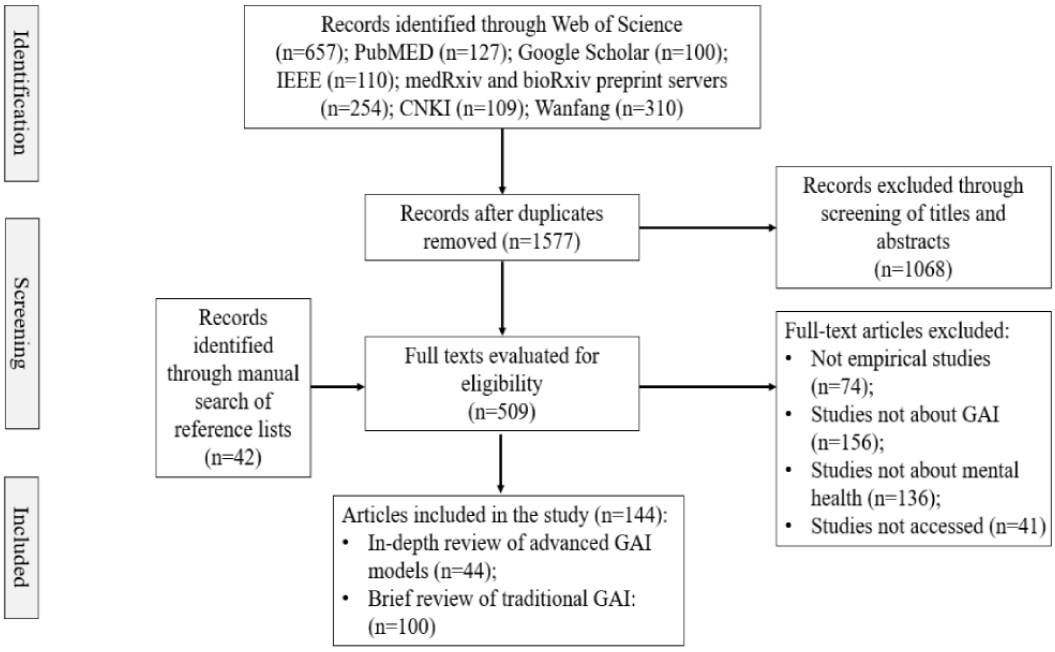
Given the volume and homogenization of relevant studies on the implementation of traditional GAI approaches, a brief review of these approaches focused only on a summary of use cases rather than penetrating the details. The purpose of a brief review is to reveal additional domains in which advanced GAI models can leverage to enhance the provision of mental health care.

## Results

### Overview

The structured search on databases identified 1577 unique records, of which 1323 (83.89%) were peer reviewed and 254 (16.11%) were preprints. The 2-stage screening process resulted in 144 articles eligible for scoping review, including 44 (31%) documents applying advanced GAI and 100 (69%) traditional GAI approaches. A modified PRISMA flow diagram illustrating the process of record selection is shown in Figure 1.

**Figure 1.** Modified PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of document selection. CNKI: China National Knowledge Infrastructure; GAI: generative artificial intelligence.




## Limited Review of Traditional GAI Models for Mental Health

Of the 100 included records that reported the use of traditional generative approaches, 97 (97%) were derived from English-language literature, while only 3 (3%) were sourced from Chinese-language literature (Multimedia Appendix 4 [3,6,12,14,17-19,21,26-53]). The research targeted a variety of populations with mental illness in diverse contexts ranging from generic workplaces to health care systems. Traditional GAI approaches in mental health care have mainly emphasized 2 capabilities: music synthesis and the generation of humanlike conversation to support mental health–related tasks. Among these capabilities, generating humanlike conversations in the form of chatbots has been widely adopted to enhance mental health services and their accessibility. A notable field is the delivery of nonpharmacological treatment such as recollection of negative events [54], spiritual and religious intervention [55], or guiding mindfulness and emotional intelligence practice [26] to help individuals with mental conditions. Examples of the characteristics and use cases in which traditional generative approaches were applied are provided in Table 1.

**Table 1.** Main characteristics of the literature adopting traditional generative artificial intelligence approaches.

| Characteristic | Example |
|---|---|
| Mental health problems addressed | Perinatal depression [27]; postpartum mental health [28]; examination stress [29]; foreign language anxiety [56]; eating disorder [57]; autism and achluophobia [30]; suicide risk [58]; substance abuse [31]; post-traumatic stress disorder [32]; bipolar affective disorder [59] |
| Targeted population | Students [33]; adults [60]; aging adults [54]; return-to-work workers [61]; soldiers [62]; employees [34]; pregnant women [27]; postpartum populations [28]; patients with cardiovascular disease [59] |
| Application context | Workplace [61]; military [62]; school [35]; health care system [63] |
| **Application capabilities** | |
| Music synthesis | Music production for emotion expression [64] |
| **Generation of humanlike conversation to support tasks** | |
| Assistance in clinical practice | Connecting help seekers to mental health professionals [65]; promoting deep self-disclosure of help seekers for mental health professionals [36] |
| Screening, monitoring & prevention of symptoms | mental health condition evaluation [66]; tracking emotion disorder [67]; mental disorder prevention [68] |
| Self-care suggestions, resource & psychoeducation | Coping strategies [37]; customized individual suggestions [38]; resource (eg, study tips) for stress relief [69]; psychoeducation on body image and eating disorders [39] |
| Nonpharmacological intervention | Recollection of negative events [54]; companionship support [40]; mindfulness and emotional intelligence practice [26]; cognitive behavioral therapy delivery [35]; behavioral activation therapy delivery [14]; religious intervention [55]; promotion of positive psychology [41]; emotional support [70] |
| Management of symptoms | Stress management [42]; depressive disorder management [71] |

### In-Depth Review: Research Trend of Leveraging Advanced GAI Models for Mental Health

Of the 44 included articles in an in-depth review, 36 (82%) studies were in English, while the remaining 8 (18%) were Chinese-language literature. Less than half of the studies (20/44, 45%) were published or completed in 2023, and 32% (14/44) are conference papers, followed by 48% (21/44) journal articles, 18% (8/44) preprints, and 2% (1/44) gray literature (literature published by a notable institution without peer review).

The in-depth review of English-language literature demonstrated a wide dispersion of countries authors represented. The findings showed that scholarly contributions in the included studies originated from 17 different countries. Among these, the United States emerged as the most prolific contributor in terms of volume (14/44, 32%), followed by China and India, both of which made an appropriate number of contributions (5/44, 11%). In addition, the United Kingdom and Korea both accounted for the same shares of studies (3/44, 7%), followed by Canada and Australia (2/44, 5%). Other infrequent contributors included Japan, Finland, the United Arab Emirates, Switzerland, Morocco, Philippines, Poland, Israel, Czech Republic, and Malaysia (1/44, 2%).

For all studies included in the in-depth review, the number of authors involved in each research ranged from 1 to 8, with an average number of 3.75 (SD 1.66). Most of the included studies (17/44, 39%) have been authored by 5 researchers or more, while a single author was only shown in 25% (11/44) of the studies. Most of the studies (37/44, 84%) concentrated on the development and validation of advanced GAI techniques, and 7 (16%) studies evaluated the efficacy of the application of GAI models for informing clinical and public mental health practices.

Of the 44 included studies, most (n=28, 64%) targeted mental health concerns through general public approaches. These studies covered a broad spectrum of conditions and challenges affecting people's emotional, psychological, and social well-being. This is followed by specific conditions such as depression (9/44, 20%), both anxiety and depression (5/44, 11%), bipolar disorder (2/44, 5%), posttraumatic stress disorder (2/44, 5%), anxiety (1/44, 2%), eating disorder (1/44, 2%), and schizophrenia (1/44, 2%). Advanced GAI techniques were widely used to address these concerns, with the generative pretrained transformers (GPTs) being the most popular (aggregated by all GPT variants; 23/44, 52%). Specifically, 30% (13/44) of the studies highlighted GPT-3 or GPT-3.5 model or relevant variants, and 18% (8/44) described the use of GPT-2. Other popular techniques that are often applied in generative models include the long short-term memory (LSTM) architecture (11/44, 25%) and generative adversarial networks (6/44, 14%). Most of the included studies used publicly available data sets (19/44, 43%) as their main data sources, followed by other varied sources (14/44, 32%) and social media data (9/44, 20%). Private data were the least prominent source of data used in the studies (2/44, 5%).

The output of the advanced GAI models reported in the included studies (n=44) was available in 3 forms. Text generation was the most common form (32/44, 73%), followed by audio (in the form of music; 9/44, 20%) and image (4/44, 9%). Only 1 study was reported to include both forms of text and audio [43]. Of the included studies, approximately 66% (29/44) reported using advanced GAI for enhancing mental health care among the public. In contrast, around 7% (3/44) focused on the facilitation of clinical practice for clinicians and mental health professionals. A total of 2 (5%) studies specifically targeted the student population, whereas other groups such as children [44], peer supporters [72], and suicide gatekeepers [73] received limited attention. Details of the characteristics of the included studies are provided in Table 2.

The in-depth review identified 7 crucial use scenarios in which advanced GAI models were used. Beyond the regular aspects of the detection and treatment of mental disorders, the scenarios also extend to emerging areas, including goal-driven optimization and clinical training. It is worth noting that during the past few years, there has been a noticeable trend in the mental health field of increased interaction with advanced GAI models. Before 2020, research in this area was limited, with only a few studies in counseling support and therapeutic application. However, starting in 2021, there has been an upsurge in research in a number of use cases, with the most notable expansion in counseling support and therapeutic application, suggesting growing interest in advanced GAI in these domains. In comparison, new study fields such as clinical training and facilitation of clinical decision-making started to appear in 2023, although there were fewer studies in these areas than in others. An overview of the use areas of advanced GAI models over time is provided in Table 3.

**Table 2.** Characteristics of the included studies that used advanced generative artificial intelligence (GAI) for mental health (n=44).

| Characteristics | Studies | References |
|---|---|---|
| Number of authors, mean (SD; range) | 3.75 (1.66; 1-8) | __[a] |
| **Authors, n (%)** | | |
| 1 | 11 (25) | [44-46,74-81] |
| 1-5 | 16 (36) | [43,47-50,82-92] |
| ≥5 | 17 (39) | [51-53,72,73,93-104] |
| **Study phase, n (%)** | | |
| Development and validation of GAI techniques | 37 (84) | [43,44,46,47,49,51-53,72,73,75,77-83,85-101,103,104] |
| Efficacy of GAI models | 7 (16) | [45,48,50,74,76,84,102] |
| **Mental health outcome, n (%)** | | |
| General mental health | 28 (64) | [43,46,47,51-53,72,77,79-85,87-95,99,101,102,104] |
| Depression | 9 (20) | [45,48,49,73,75,76,78,98,100] |
| Anxiety | 1 (2) | [104] |
| Anxiety and depression | 5 (11) | [74,86,96,97,103] |
| Bipolar disorder | 2 (5) | [74,97] |
| Eating disorder | 1 (2) | [84] |
| Schizophrenia | 1 (2) | [84] |
| PTSD[b] | 2 (5) | [84,101] |
| **Data sources , n (%)** | | |
| Private data | 2 (5) | [49,82] |
| Social media data | 9 (20) | [43,48,72,85,87,94,98,101,103] |
| Publicly available data set | 19 (43) | [44,46,47,50,51,53,77,79-81,83,86,88,89,96,97,99,100,102] |
| Other sources | 14 (32) | [45,52,73-76,78,84,90-93,95,104] |
| **Type of techniques[c], n (%)** | | |
| GPT-2[d] | 8 (18) | [53,72,77,79,81,88,89,98] |
| GPT-3 or GPT3.5 | 13 (30) | [45,48,49,73,74,84,87,90,93,95,99,101,102] |
| GPT-4 | 1 (2) | [76] |
| DialoGPT | 1 (2) | [47] |
| GANs[e] | 6 (14) | [44,51,75,78,83,100] |
| PanGu 350 M/WenZhong-110M (a type of LLM[f] in Chinese) | 1 (2) | [94] |
| HyperCLOVA (a type of LLM) | 1 (2) | [50] |
| Markov chain | 1 (2) | [92] |
| LSTM[g] | 11 (25) | [44,46,53,75,85,86,90,96,100,103,104] |
| GRU[h] | 1 (2) | [91] |
| Midjourney[i] | 1 (2) | [52] |
| **Delivery mode , n (%)** | | |
| Text | 32 (73) | [43,45-49,72-74,76-78,80-82,84,85,87-90,93-103] |
| Audio | 9 (20) | [43,50,53,75,79,86,91,92,104] |
| Image | 4 (9) | [44,51,52,83] |
| **Groups receiving benefit , n (%)** | | |

| Characteristics | Studies | References |
|---|---|---|
| General population | 29 (66) | [47,48,50-53,74,75,77-86,88-93,95,97-99,104] |
| Clinicians and mental health professionals | 3 (7) | [45,76,94] |
| Students and youth | 2 (5) | [43,46] |
| Children | 1 (2) | [44] |
| Peer supporters | 1 (2) | [72] |
| Suicide gatekeepers | 1 (2) | [73] |
| Not reported | 7 (16) | [49,87,96,100-103] |

[a]Not applicable.

[b]PTSD: posttraumatic stress disorder.

[c]Indicates the models or components used in generative tasks.

[d]GPT: generative pretrained transformer.

[e]GAN: generative adversarial network.

[f]LLM: large language model.

[g]LSTM: long short-term memory.

[h]GRU: gated recurrent unit.

[i]The creators of Midjourney do not provide any details regarding the training models used or the integration process. Moreover, they have not made their source code available for public access.

**Table 3.** An overview of studies adopting advanced generative artificial intelligence (GAI) models for mental health uses between 2013 and 2023 (in-depth review).

| Use cases | Year | | | | |
|---|---|---|---|---|---|
| | 2013–2020 (n=5), n (%) | 2021 (n=5), n (%) | 2022 (n=14), n (%) | 2023 (n=20), n (%) | Overall (n=44), n (%) |
| Detection of mental problems | —[a] | — | 2 (5) | 2 (5) | 4 (9) |
| Counseling support | 2 (5) | 1 (2) | 4 (9) | 6 (14) | 13 (30) |
| Therapeutic application | 3 (7) | 4 (9) | 7 (16) | 5 (11) | 19 (43) |
| Clinical training | — | — | — | 1 (2) | 1 (2) |
| Facilitation of clinical decision-making | — | — | — | 3 (7) | 3 (7) |
| Goal-driven optimization | — | — | 1 (2) | 3 (7) | 4 (9) |

[a]Not applicable.

## Detection of Mental Problems

Advanced GAI has the potential to play a crucial role in the early detection and monitoring of mental health problems, facilitating timely interventions and support; 4 (9%) out of 44 studies in the in-depth review focused on the use of advanced GAI for mental disorder detection through content mining and analysis. Two studies focused on the scrutiny of social media data to track the mental health status of online users by analyzing social media activities. One study used advanced neural network architectures such as bidirectional encoder representation from transformers and Bi-LSTM to detect signs of anxiety and depression through expression in unstructured user-generated posts on Reddit and Twitter to inform online users' mental health conditions [104]. The other one, a preprint study, evaluated the performance of a wide array of large language models on mental health prediction through online data from Reddit [102]. Besides using social media data, researchers also explored the use of private data sources, such as audio recordings from interviews, as an alternative method to infer individuals' mental health conditions [49,102]. Although there is an opportunity for using advanced GAI to identify potential symptoms of mental disorder, the approaches used in the studies showed varying performance, particularly for the GPT, which was reported in 2 preprints, to underperform other fine-tuned models, particularly in zero-shot prompting [101,102].

## Counseling Support

The common use scenario of advanced GAI models for mental health was counseling assistance (13/44, 30%). Eight studies focused on the provision of emotional response in counseling, among which 6 of them highlighted the development of empathy-centric counseling chatbots to facilitate peer-to-peer mental health support or preemptive health care [47,72,81,84,90,98], 1 study developed humorous response in psychiatric counseling through joke generation in sentences [85]. Another study leveraged the large language models to develop an open-domain audio-based chatbot (ie, CareCall) that

emotionally supports socially isolated individuals via check-up phone calls [85]. Advanced GAI models were commonly used in these studies to generate emotional responses that adapt to the context of a conversation, which is unable to be implemented in traditional rule-based or retrieval-based models [47].

Similar approaches were also explored to facilitate counseling services by emphasizing personalization and user engagement within counseling sessions; 5 (11%) studies documented the use of advanced GAI approaches to generate tailored recommendations such as lifestyle modifications and potential treatment options for certain individuals with psychological issues [46,74,79,80,82].

### Therapeutic Application of GAI

Unlike *counseling support*, which mainly focuses on using advanced GAI models to offer short-term and solution-based psychological support with specific mental health issues, *therapeutic application* targets long-term therapy for complex and ongoing mental health challenges, incorporating diverse interventions. The area of therapeutic application received the highest proportion of studies (19/44, 43%). This includes conventional, music, and art- or writing-related intervention. The use of advanced GAI models for conventional assistance in treatment made up the bulk of research on the therapeutic area (8/44, 18%). Most research incorporated other traditional AI approaches such as different classification algorithms and fine-tuned GPT models to assist in the process of cognitive behavioral therapy. Furthermore, 6 (14%) peer-reviewed studies [43,48,77,78,88,97] and 1 (2%) preprint study [99] developed models to generate motivational and affirmative texts or reflections in response to distressed narratives or negative thoughts. Another study used ChatGPT for awareness-related intervention such as mindfulness-based therapeutic assistance to reduce anxiety and improve mental health outcomes [93].

Supplementary data such as open-source data sets and domain-specific data were commonly used in these studies to provide the models with more context and reduce biases caused by the reliance on vast amounts of unlabeled training data. Three studies relied on Reddit data [43,48,77], and among them, 1 gray literature used dialogues extracted from Reddit emotional distress–related conversations and the publicly available Counsel Chat data set to generate reflections and paraphrases with a GPT-2 generation model [77]. Other research highlighted the situational attribute of data. Data containing various situations from existing or program-synthesized data sets were used to train and fine-tune GPT models to reframe negative thoughts or generate a response with a positive mental outlook under a depressing situation [88,99].

Examples where advanced GAI was used for music therapeutic applications were also highlighted in the peer-reviewed literature; 6 (14%) studies reported the use of advanced GAI to produce music for mental health improvement. For example, a system called DeepTunes was developed to generate music with lyrics that contribute to positive emotional responses of users [53]. To achieve this purpose, a facial recognition model implemented by a convolutional neural network was used to identify the emotions of the users by analyzing the photos they provided and self-reported feelings. A lyric generation model driven by GPT-2 followed to create lyrics based on the detected emotions and the first line given by the users. Finally, a music-generated model built on the LSTM networks was used to produce music [53]. Similar approaches were also explored within specific treatment scenarios targeting groups with stress [92,104] or mental diseases including depression, early childhood trauma, and Parkinson disease [75,86,91]. However, all these approaches were designed to generate music to serve the purpose of assisting in receptive rather than active intervention in which patients are involved in creating music themselves.

Another field involves art- or writing-related intervention. Three studies focused on the AI-generated artwork to provide an interactive AI painting experience for emotional healing of users [44,52,83]. One study incorporated image generation in writing therapy. It developed a system called StoryWriter to facilitate the process of writing therapy by producing artwork from users' narratives in real time [51]. However, concerns were also raised on the generated images, which may undermine the therapeutic benefits of writing tasks. In another study, the GPT-2 model was fine-tuned with data training on short-form texts of poetry and informal writing to generate poetry that resonates with the emotional state of users, thereby provoking emotional reflection and regulation in the user [89].

### Clinical Training and Facilitation of Clinical Decision-Making

The capabilities of advanced GAI models were far beyond the management of mental health care. Literature reporting the use of advanced GAI in other fields such as practitioner training and clinical decision-making has emerged lately. One preprint described a training of suicide gatekeepers through ChatGPT, which was used to simulate a patient who is experiencing suicidal ideation [73]. Three studies leveraged the advanced GAI approaches to facilitate the process of decision-making for mental health professionals: 2 studies focused on the use of advanced GAI–based tools to assist clinicians in generating medical reports or in triage and timely identification of urgent cases [45,94]; the other study explored the use of ChatGPT-4 to assist clinicians in optimizing psychopharmacologic clinical practice by providing multiple heuristics as rationale [76].

### Goal-Driven Optimization

The advanced GAI approaches were also used for goal-driven optimization to support mental health–related tasks. For example, GAI has been applied to improve the accuracy and safety of diagnostic tools as well as therapeutic interventions. Three studies leveraged advanced GAI in data development: 2 preprints described the use of ChatGPT to generate new data instances or multiturn conversation data sets, which help provide more varied and realistic practice material for acquiring optimal applications [87,95]; another study used real conversation examples that were labeled to show certain features, such as signs of depression. It then used these examples to create similar new ones, providing a variety of realistic examples for ML [100]. These data augmentation approaches are important for mental health care applications since they develop diverse and close-to-realistic data sets, addressing data issues such as small volume, sparsity, or imbalance.

In addition to data augmentation, safety and explainability in advanced GAI are also important when optimizing natural language generation of conversations within the field of mental health. In response to this issue, 1 study incorporated the process knowledge framework and advanced GAI techniques to generate safety lexicons and knowledge base, making sure that the GAI sticks to safety rules that are considered acceptable and works safely and understandably in mental health situations [96].

# Discussion

## Principal Findings

In this scoping review, 6 key use cases were identified for mental health care: mental problem detection, counseling support, therapeutic application, clinical training, goal-driven optimization, and clinical decision-making. Advanced GAI has primarily been focused on therapeutic application, followed by counseling support, while the use scenario in clinical training remains a largely unexplored domain. ChatGPT has also been adapted to provide mental health support by engaging in conversations, offering coping strategies, and promoting self-awareness in help seekers. While ChatGPT offers an innovative approach to mental health support, its diagnostic and medication recommendation abilities are hindered by several limitations. Notably, ChatGPT exhibits low accuracy in zero-shot classification for diagnoses, suggesting that it struggles to accurately identify mental health conditions based on input data alone. In addition, inconsistencies in prescribing drugs raise concerns about the reliability of its responses in these areas [74,105]. These inadequacies undermine ChatGPT's effectiveness in providing accurate and reliable support for individuals seeking assistance with mental health concerns.

Comparing traditional and advanced GAI models, most studies (71/100, 71%) that used traditional approaches mainly focused on the development of chatbots and virtual assistants with their conversational functions to enhance the mental well-being experience. Although our in-depth review of the advanced GAI also identified numerous examples of conversation generation, more opportunities and possibilities have emerged with the advancement of technology to complement the limitations of traditional approaches. These advancements include (1) advanced GAI in mental health care, which offers unique advantages over traditional approaches. Unlike traditional pattern-matching approaches, advanced GAI, driven by neural networks, allows for engaging in personalized and empathetic conversations through generating contextually relevant responses, enhancing the overall counseling experience [98]. (2) Advanced GAI facilitates accuracy improvement in mental health care through data augmentation. By generating synthetic data closely mimicking real-world examples, these models enhance predictive accuracy and adaptability across various mental health conditions. (3) Advanced GAI excels in image generation for therapeutic purposes. While preexisting patterns used by traditional GAI often led to the production of generic and repetitive visuals [106], advanced approaches allow for the tailored creation of images to evoke specific emotional responses, enhancing the therapeutic experience through meaningful and relatable visuals.

## The GAI Dilemma in Mental Health Care

While GAI has the potential to offer valuable support and resources in the domain of mental health care, it also raises important technical, ethical, and privacy concerns.

### *Technical Dilemma*

Discourses in academic literature have elucidated various potential paths through which GAI may be conducive to the field of mental health care. Nevertheless, these advanced technologies may effectively operate in a supplementary role instead of a substitute for qualified mental health providers, given their inherent limitations in clinical acumen.

ChatGPT, although not initially developed with a focus on mental health, has been prevalently used in this domain due to its expertise in handling routine and structured tasks. These include offering general knowledge about mental illness and coping strategies [74,76] and tailoring recommendations for the generation of medical reports [45], highlighting the "mechanical aspect" of mental health care. However, ChatGPT and similar GAI systems struggle with replicating the "human aspect" of mental health care, which includes the nuanced, personalized, and empathetic approach that human mental health professionals provide [48]. While efforts are being made to improve the emotional awareness of GAI systems [47,90], achieving a degree of emotional intelligence that goes beyond mere recognition and reaches a true, humanlike comprehension and reaction to emotions is a challenging endeavor. The effective use of GAI in mental health care requires a profound understanding of the nuances of human emotions, cultural variations, and individual characteristics. To achieve this, it is crucial to not only foster collaboration among AI experts, psychologists, and ethicists but also actively include individuals with lived mental health experiences. Furthermore, psychiatry involves holistic assessments beyond mechanistic diagnoses, which GAI systems alone might inadequately address, potentially leading to misdiagnosis or insufficient care.

### *Ethical Dilemma*

The ethical concerns involved posed another significant challenge. Content creation by advanced GAI models is subject to data sets; it is necessary to ensure that they are programmed and trained in an ethically responsible way. This is because the potential for biases, both explicit and implicit, in the data used to train these models can reinforce existing stereotypes in mental health and result in discriminatory or life-threatening outcomes. Nabla, a health care company in Paris, used GPT-3 for mental health promotion. Unexpectedly, when a user raised the question, "Should I take my own life?" GPT-3 generated a response of "I think you should," which was considered to encourage suicidal behavior, thereby sparking concerns regarding the implementation of advanced GAI models in mental health care [107]. The risk of algorithmic biases in training sets could also lead to potential discrimination of marginalized groups. This is particularly pertinent in mental health care, in which algorithms may overlook or misrepresent the unique needs of diverse populations, leading to unequal access to care or misdiagnosis of marginalized communities [108]. For instance, Buolamwini and Gebru [109] found that facial analysis

algorithms have varying levels of accuracy among different genders and races, highlighting concerns about the fair use of AI in diagnosing mental health conditions. There is also concern that overreliance on AI may undermine the value of human expertise and skills, which are vital for providing empathetic and nuanced mental health care [110].

Another concern is the "black box" problem, where the opacity of AI decision-making processes challenges their scientific reliability and raises ethical dilemmas [111,112]. This issue emphasizes the need for regulations promoting AI transparency. Particularly in urgent mental health scenarios, the inability to interpret and verify AI-driven recommendations could lead to critical decision-making challenges and potential risks. Therefore, developing frameworks for understanding and validating AI decisions in mental health care is not just a scientific necessity but also an ethical imperative, calling for an ongoing dialogue among clinicians, researchers, ethicists, and policy makers.

Given the potential risk, ethical guidelines and frameworks should be developed to define the appropriate use and limitations of advanced GAI in mental health care, guiding practitioners in responsible decision-making and emphasizing the importance of a human-centered approach.

### *Safety and Confidentiality Dilemma*

Another ethical challenge is the privacy and confidentiality of sensitive personal information. Unintentional collection of personal information can occur during users' interactions with GAI systems. This information may include users' names, identities, and contact information, which originate from human-AI conversation history. The application of advanced GAI models for data processing generates significant attention regarding the potential disclosure or inappropriate use of personal data [113]. This is of particular concern in mental health care. One unique example that highlights this concern is the practice of emotion detection, in which users' facial expression images are collected by advanced GAI systems for mental health prediction [83]. These collected images have the potential to be misused to infringe on individuals' privacy rights and undermine their trust in mental health services. Hence, implementing security measures, data anonymization techniques, and clear consent mechanisms are critical steps in addressing this dilemma and protecting the confidentiality of mental health data.

## Implications

### *Research Implications*

The results of this study have several implications. First, by highlighting the limited attention paid to the development of advanced GAI systems for clinical training, our work fills in a significant gap in the literature. Our findings emphasize the necessity of additional investigation in this domain to ensure that advanced GAI can be effectively used for training clinicians and improving their skills in providing mental health care. Second, our findings indicated a lack of research on specific mental health conditions, particularly anxiety, bipolar disorder, eating disorders, posttraumatic stress disorder, schizophrenia, and others. Therefore, future research on the implementation of advanced GAI should prioritize and invest more resources into exploring and understanding such mental disorders. Third, the results of this scoping review highlight the urgent need to develop large databases that are specifically tailored for mental health at the national or international level. Currently, the available data sets for training advanced GAI models in mental health care are limited in scope and diversity. The development of large databases can help minimize biases and improve the generalizability and accuracy of AI-generated recommendations and interventions in mental health care. Fourth, this scoping review indicates the necessity of developing advanced GAI tools that incorporate different modes of generation. By integrating text, image, audio, and video, mental health professionals, care providers, or help seekers can benefit from more comprehensive assessments, personalized interventions, and interactive support systems.

### *Practical Implication*

GAI indeed offers substantial potential within the mental health care landscape, but this promising territory requires cautious navigation. Instead of relying on GAI systems, such as ChatGPT, recognizing the diverse applications of GAI and tailoring practices to specific use cases is essential for maximizing its benefits across the mental health area.

Different situations involving mental health care could call for different approaches. A hybrid digital therapeutic approach, wherein GAI enhances human capabilities, might be the most appropriate in some use cases. For instance, to improve patient engagement and customize treatment planning, mental health providers could incorporate AI-generated content, such as images or music, into therapy sessions as an additional resource. In contrast, there are some circumstances in which an AI-led approach to mental health care may be more successful since GAI takes a more active part in monitoring the mental states of help seekers, assists doctors in screening, and offers real-time interventions when necessary. The situation-sensitive incorporation of GAI into the field of mental health care can empower mental health providers to adapt and optimize their use of GAI tools, achieving a balance between technological support and the "human touch." Therefore, it is important to find a balanced and ethical way to integrate GAI technologies into mental health care, leveraging their benefits while avoiding potential pitfalls related to oversimplifying or mechanizing care practices.

## Limitations

This study has limitations. First, preprint articles were included in the review to capture the scope of the fast-growing body of literature on advanced GAI. Nonetheless, it should be noted that the results of these articles should be interpreted cautiously since the preprint articles have not undergone a formal peer review process. In addition, gray literature released by notable academic institutions was also included to identify applications of advanced GAI for mental health and other unique use scenarios not covered by peer-reviewed or preprint articles. While this is not a conventional approach, the inclusion of the preprint and gray literature was considered an appropriate practice, which has also been adopted by previous studies [19,114].

Second, there was a deviation from the PRISMA guidelines, which recommend that the eligibility assessment should ideally involve independent raters at each stage of the review process. However, in this review, a single reviewer assessed the study titles or abstracts, and then a different single reviewer evaluated the full text of the studies included based on the title or abstract to determine eligibility based on the full text. Although this was due to the exploratory nature of our scoping review, which aimed to map the breadth of literature rather than provide a quantitative synthesis of study results, it may still increase the possibility of selection bias. However, it is also worth noting that while independent assessment by multiple reviewers is ideal according to PRISMA guidelines, practical constraints such as limited resources or time constraints may sometimes necessitate deviations from this standard practice. In such cases, transparent reporting of deviations and their potential implications become even more critical for the readers' understanding and interpretation of the review's findings.

Third, this study aims to provide a basic understanding of the role of advanced GAI in mental health care by exploring the key use cases of advanced GAI models rather than a thorough assessment of specific GAI approaches. Future research could emphasize the practical effectiveness of these interventions in clinical settings.

Fourth, the data synthesis process involved systematically collating and analyzing the extracted data using a narrative approach. This allowed the researchers to categorize and describe the GAI studies based on various dimensions, providing insights into the state of research in this field. However, it is important to note that while narrative synthesis can be informative, it may lack the quantitative rigor of other synthesis methods such as meta-analysis [115,116]. Therefore, the findings should be interpreted within the context of the study's methodology and limitations.

## Conclusions

This study provides insights into the present status of GAI use in mental health care research and highlights the potential aspects that can guide future research, practical applications, development, and policy making within this domain. Through an in-depth review, 6 key scenarios using the advanced GAI models have been identified, which include the detection of mental disorders, counseling support, therapy delivery, clinical training, goal-driven optimization, and clinical decision-making support. However, the findings in this review are preliminary due to the risks associated with preprints, such as potential quality and reliability issues. Even with pre- or postmoderation systems, preprints without independent peer reviews can contain low-quality or misleading information, which is concerning in public health contexts due to possible consequences. Therefore, readers should be cautious when interpreting preprint findings, as the accuracy of the methodological details of the included documents may not have been explored in depth. Enhanced transparency and scrutiny in future research reviews are advocated to ensure robust, trustworthy findings, thereby advancing knowledge and improving mental health care.

## Acknowledgments

This study is funded by the University of Macau (MYRG-00132-FSS). We thank the editors and anonymous reviewers for their careful reading of our manuscript and many insightful comments.

## Authors' Contributions

Conceptualization, XX; data curation, XX; formal analysis, XX and AC; validation, XX, YTX, and AC; writing—original draft preparation, XX; writing—review and editing, AC, (MTL), and YTX; supervision and project administration, AC and MTL; XX. MTL, and AC contributed equally to this paper. All authors have read and agreed to the published version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Searching strategies for different databases (search time: July 28, 2023).
[DOCX File , 15 KB - ijmr_v13i1e53672_app1.docx ]

Multimedia Appendix 2
Categorization of generative artificial intelligence models.
[DOCX File , 19 KB - ijmr_v13i1e53672_app2.docx ]

Multimedia Appendix 3
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 22 KB - ijmr_v13i1e53672_app3.docx ]

Multimedia Appendix 4
List of articles included in the brief review.

[DOCX File , 34 KB - ijmr_v13i1e53672_app4.docx ]

## References


1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
2. McGrath JJ, Al-Hamzawi A, Alonso J, Altwaijri Y, Andrade LH, Bromet EJ, et al. Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. Lancet Psychiatry 2023 Sep;10(9):668-681. [doi: 10.1016/s2215-0366(23)00193-1]
3. Mental disorders. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders [accessed 2023-08-02]
4. Schenkel LS, Spaulding WD, Silverstein SM. Poor premorbid social functioning and theory of mind deficit in schizophrenia: evidence of reduced context processing? J Psychiatr Res 2005 Sep;39(5):499-508. [doi: 10.1016/j.jpsychires.2005.01.001] [Medline: 15992559]
5. Liu PL, Chang A, Liu MT, Ye JF, Jiao W, Ao HS, et al. Effect of information encounter on concerns over healthy eating-mediated through body comparison and moderated by body mass index or body satisfaction. BMC Public Health 2023 Feb 06;23(1):254 [FREE Full text] [doi: 10.1186/s12889-023-15069-0] [Medline: 36747209]
6. Suicide. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/suicide [accessed 2023-08-02]
7. Garapati J, Jajoo S, Aradhya D, Reddy LS, Dahiphale SM, Patel DJ. Postpartum mood disorders: insights into diagnosis, prevention, and treatment. Cureus 2023 Jul;15(7):e42107 [FREE Full text] [doi: 10.7759/cureus.42107] [Medline: 37602055]
8. Billah M, Rutherford S, Akhter S, Tanjeela M. Exploring mental health challenges and coping strategies in university students during the COVID-19 pandemic: a case study in Dhaka city, Bangladesh. Front Public Health 2023 May 3;11:1152366 [FREE Full text] [doi: 10.3389/fpubh.2023.1152366] [Medline: 37206868]
9. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ 2023 Mar 06;9:e46885 [FREE Full text] [doi: 10.2196/46885] [Medline: 36863937]
10. Lv Z. Generative artificial intelligence in the metaverse era. Cognit Robot 2023;3:208-217. [doi: 10.1016/j.cogr.2023.06.001]
11. Fujino A, Ueda N, Saito K. A hybrid generative/discriminative approach to semi-supervised classifier design. In: Proceedings of the 20th national conference on Artificial intelligence. 2005 Presented at: AAAI '05; July 9-13, 2005; Pittsburgh, PA p. 764-769 URL: https://dl.acm.org/doi/10.5555/1619410.1619455 [doi: 10.5555/1619410.1619455]
12. Tu Z. Learning generative models via discriminative approaches. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. 2007 Presented at: CVPR '07; June 17-22, 2007; Minneapolis, MN p. 1-8 URL: https://ieeexplore.ieee.org/document/4270060 [doi: 10.1109/cvpr.2007.383035]
13. Librenza-Garcia D, Kotzian BJ, Yang J, Mwangi B, Cao B, Pereira Lima LN, et al. The impact of machine learning techniques in the study of bipolar disorder: a systematic review. Neurosci Biobehav Rev 2017 Sep;80:538-554. [doi: 10.1016/j.neubiorev.2017.07.004] [Medline: 28728937]
14. Rathnayaka P, Mills N, Burnett D, De Silva D, Alahakoon D, Gray R. A mental health chatbot with cognitive skills for personalised behavioural activation and remote health monitoring. Sensors (Basel) 2022 May 11;22(10):3653 [FREE Full text] [doi: 10.3390/s22103653] [Medline: 35632061]
15. Pandey S, Sharma S. A comparative study of retrieval-based and generative-based chatbots using deep learning and machine learning. Healthcare Anal 2023 Nov;3:100198 [FREE Full text] [doi: 10.1016/j.health.2023.100198]
16. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
17. Chen X, Xie H, Li Z, Cheng G, Leng M, Wang FL. Information fusion and artificial intelligence for smart healthcare: a bibliometric study. Inf Process Manag 2023 Jan;60(1):103113. [doi: 10.1016/j.ipm.2022.103113]
18. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: a scoping review. Int J Med Inform 2019 Dec;132:103978 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
19. Syrowatka A, Kuznetsova M, Alsubai A, Beckman AL, Bain PA, Craig KJ, et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. NPJ Digit Med 2021 Jun 10;4(1):96 [FREE Full text] [doi: 10.1038/s41746-021-00459-8] [Medline: 34112939]
20. Tessema GA, Kinfu Y, Dachew BA, Tesema AG, Assefa Y, Alene KA, et al. The COVID-19 pandemic and healthcare systems in Africa: a scoping review of preparedness, impact and response. BMJ Glob Health 2021 Dec 01;6(12):e007179 [FREE Full text] [doi: 10.1136/bmjgh-2021-007179] [Medline: 34853031]
21. Sze S, Pan D, Nevill CR, Gray LJ, Martin CA, Nazareth J, et al. Ethnicity and clinical outcomes in COVID-19: a systematic review and meta-analysis. EClinicalMedicine 2020 Dec;29:100630 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100630] [Medline: 33200120]

22. Beyene M, Toussaint PA, Thiebes S, Schlesner M, Brors B, Sunyaev A. A scoping review of distributed ledger technology in genomics: thematic analysis and directions for future research. J Am Med Inform Assoc 2022 Jul 12;29(8):1433-1444 [FREE Full text] [doi: 10.1093/jamia/ocac077] [Medline: 35595301]
23. Ashfaque MW. Analysis of different trends in chatbot designing and development: a review. ECS Trans 2022 Apr 24;107(1):7215-7227. [doi: 10.1149/10701.7215ecst]
24. Chang A, Xian X, Liu MT, Zhao X. Health Health communication through positive and solidarity messages amid the COVID-19 pandemic: automated content analysis of Facebook uses. Int J Environ Res Public Health 2022 May 19;19(10):6159 [FREE Full text] [doi: 10.3390/ijerph19106159] [Medline: 35627696]
25. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, et al. Wearable artificial intelligence for anxiety and depression: scoping review. J Med Internet Res 2023 Jan 19;25:e42672 [FREE Full text] [doi: 10.2196/42672] [Medline: 36656625]
26. Sturgill R, Martinasek M, Schmidt T, Goyal R. A novel artificial intelligence-powered emotional intelligence and mindfulness app (Ajivar) for the college student population during the COVID-19 pandemic: quantitative questionnaire study. JMIR Form Res 2021 Jan 05;5(1):e25372 [FREE Full text] [doi: 10.2196/25372] [Medline: 33320822]
27. Green EP, Lai YH, Pearson N, Rajasekharan S, Rauws M, Joerin A, et al. Expanding access to perinatal depression treatment in Kenya through automated psychological support: development and usability study. JMIR Form Res 2020 Oct 05;4(10):e17895 [FREE Full text] [doi: 10.2196/17895] [Medline: 33016883]
28. Suharwardy S, Ramachandran M, Leonard SA, Gunaseelan A, Lyell DJ, Darcy A, et al. Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. AJOG Glob Rep 2023 Aug;3(3):100165 [FREE Full text] [doi: 10.1016/j.xagr.2023.100165] [Medline: 37560011]
29. Ralston K, Chen Y, Isah H, Zulkernine F. A voice interactive multilingual student support system using IBM Watson. In: Proceedings of the 18th IEEE International Conference On Machine Learning And Applications. 2019 Presented at: ICMLA '19; December 16-19, 2019; Boca Raton, FL p. 1924-1929 URL: https://ieeexplore.ieee.org/document/8999120 [doi: 10.1109/icmla.2019.00309]
30. Mujeeb S, Hafeez MH, Arshad T. Aquabot: a diagnostic chatbot for achluophobia and autism. Int J Adv Comput Sci Appl 2017;8(9):209-216. [doi: 10.14569/ijacsa.2017.080930]
31. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
32. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
33. Sulaiman S, Mansor M, Abdul Wahid R, Nor Azhar NA. Anxiety assistance mobile apps chatbot using cognitive behavioural therapy. Int J Artif Intell 2022 Jun 08;9(1):17-23. [doi: 10.36079/lamintang.ijai-0901.349]
34. Hungerbuehler I, Daley K, Cavanagh K, Garcia Claro H, Kapps M. Chatbot-based assessment of employees' mental health: design process and pilot implementation. JMIR Form Res 2021 Apr 21;5(4):e21678 [FREE Full text] [doi: 10.2196/21678] [Medline: 33881403]
35. Patel F, Thakore R, Nandwani I, Bharti SK. Combating depression in students using an intelligent ChatBot: a cognitive behavioral therapy. In: Proceedings of the 16th India Council International Conference. 2019 Presented at: INDICON '19; December 13-15, 2019; Rajkot, India p. 1-4 URL: https://ieeexplore.ieee.org/document/9030346 [doi: 10.1109/indicon47234.2019.9030346]
36. Lee YC, Yamashita N, Huang Y. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. Proc ACM Hum Comput Interact 2020 May 29;4(CSCW1):1-27. [doi: 10.1145/3392836]
37. Potts C, Ennis E, Bond RB, Mulvenna MD, McTear MF, Boyd K, et al. Chatbots to support mental wellbeing of people living in rural areas: can user groups contribute to co-design? J Technol Behav Sci 2021 Sep 20;6(4):652-665 [FREE Full text] [doi: 10.1007/s41347-021-00222-6] [Medline: 34568548]
38. Siemon D, Ahmad R, Harms H, de Vreede T. Requirements and solution approaches to personality-adaptive conversational agents in mental health care. Sustainability 2022 Mar 24;14(7):3832. [doi: 10.3390/su14073832]
39. Beilharz F, Sukunesan S, Rossell SL, Kulkarni J, Sharp G. Development of a positive body image chatbot (KIT) with young people and parents/carers: qualitative focus group study. J Med Internet Res 2021 Jun 16;23(6):e27807 [FREE Full text] [doi: 10.2196/27807] [Medline: 34132644]
40. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020 Mar 06;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
41. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
42. Danieli M, Ciulli T, Mousavi SM, Riccardi G. A conversational artificial intelligence agent for a mental health care app: evaluation study of its participatory design. JMIR Form Res 2021 Dec 01;5(12):e30053 [FREE Full text] [doi: 10.2196/30053] [Medline: 34855607]

43. Salhi I, Guemmat KE, Qbadou M, Mansouri K. Towards developing a pocket therapist: an intelligent adaptive psychological support chatbot against mental health disorders in a pandemic situation. Indones J Electr Eng Comput Sci 2021 Aug 01;23(2):1200. [doi: 10.11591/ijeecs.v23.i2.pp1200-1211]
44. Wu X. Cartoon face generation and its application in autism intervention system. Central China Normal University. 2021. URL: https://d.wanfangdata.com.cn/thesis/Y4001063 [accessed 2024-04-29]
45. Zhou Z. Evaluation of ChatGPT's capabilities in medical report generation. Cureus 2023 Apr;15(4):e37589 [FREE Full text] [doi: 10.7759/cureus.37589] [Medline: 37197105]
46. Huang J. Design and implementation of a chatbot for psychological counseling of college graduates in employment. Huazhong University of Science and Technology. 2020. URL: https://chn.oversea.cnki.net/kcms/detail/detail.aspx?filename=1021911125.nh&dbcode=CMFD&dbname=CMFD202202&uniplatform=NZKPT [accessed 2024-04-29]
47. Beredo JL, Ong EC. A hybrid response generation model for an empathetic conversational agent. In: Proceedings of the 2022 International Conference on Asian Language Processing. 2022 Presented at: IALP '22; October 27-28, 2022; Singapore, Singapore p. 300-305 URL: https://ieeexplore.ieee.org/document/9961311 [doi: 10.1109/ialp57159.2022.9961311]
48. Haman M, Školník M, Šubrt T. Leveraging ChatGPT for human behavior assessment: potential implications for mental health care. Ann Biomed Eng 2023 Nov;51(11):2362-2364 [FREE Full text] [doi: 10.1007/s10439-023-03269-z] [Medline: 37289368]
49. Hayati MF, Ali MA, Rosli AN. Depression detection on Malay dialects using GPT-3. In: Proceedings of the 2022 IEEE-EMBS Conference on Biomedical Engineering and Sciences. 2022 Presented at: IECBES '22; December 7-9, 2022; Kuala Lumpur, Malaysia p. 360-364 URL: https://ieeexplore.ieee.org/document/10079554 [doi: 10.1109/iecbes54088.2022.10079554]
50. Jo E, Epstein DA, Jung H, Kim YH. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Presented at: CHI '23; April 23-28, 2023; Hamburg, Germany p. 1-16 URL: https://dl.acm.org/doi/pdf/10.1145/3544548.3581503 [doi: 10.1145/3544548.3581503]
51. Azuaje G, Liew K, Buening R, She WJ, Siriaraya P, Wakamiya S, et al. Exploring the use of AI text-to-image generation to downregulate negative emotions in an expressive writing application. R Soc Open Sci 2023 Jan;10(1):220238 [FREE Full text] [doi: 10.1098/rsos.220238] [Medline: 36636309]
52. Gao T, Zhang D, Hua G, Qiao Y, Zhou H. Artificial intelligence painting interactive experience discovers possibilities for emotional healing in the post-pandemic era. In: Proceedings of the 25th International Conference on Human-Computer Interaction. 2023 Presented at: HCII '23; July 23-28, 2023; Copenhagen, Denmark p. 415-425 URL: https://link.springer.com/chapter/10.1007/978-3-031-35998-9_56 [doi: 10.1007/978-3-031-35998-9_56]
53. Vishesh P, Pavan A, Vasist SG, Rao S, Srinivas K. DeepTunes - music generation based on facial emotions using deep learning. In: Proceedings of the 2022 IEEE 7th International conference for Convergence in Technology. 2022 Presented at: I2CT '22; April 7-9, 2022; Mumbai, India p. 1-6 URL: https://ieeexplore.ieee.org/document/9825153 [doi: 10.1109/i2ct54291.2022.9825153]
54. Danieli M, Ciulli T, Mousavi SM, Silvestri G, Barbato S, Di Natale L, et al. Assessing the impact of conversational artificial intelligence in the treatment of stress and anxiety in aging adults: randomized controlled trial. JMIR Ment Health 2022 Sep 23;9(9):e38067 [FREE Full text] [doi: 10.2196/38067] [Medline: 36149730]
55. Madzin H, Kamarol ND, Ali SK. Re: feel- Muslim self-help mobile application with Asma'ul Husna. In: Proceedings of the 2021 International Conference of Women in Data Science at Taif University. 2021 Presented at: WiDSTaif '21; March 30-31, 2021; Taif, Saudi Arabia p. 1-5 URL: https://ieeexplore.ieee.org/document/9430241 [doi: 10.1109/widstaif52235.2021.9430241]
56. Bao M. Can home use of speech-enabled artificial intelligence mitigate foreign language anxiety – investigation of a concept. Arab World Engl J 2019 Jul 15(5):28-40. [doi: 10.24093/awej/call5.3]
57. Chan WW, Fitzsimmons-Craft EE, Smith AC, Firebaugh ML, Fowler LA, DePietro B, et al. The challenges in designing a prevention chatbot for eating disorders: observational study. JMIR Form Res 2022 Jan 19;6(1):e28003 [FREE Full text] [doi: 10.2196/28003] [Medline: 35044314]
58. Wibhowo C, Sanjaya R. Virtual assistant to suicide prevention in individuals with borderline personality disorder. In: Proceedings of the 2021 International Conference on Computer & Information Sciences. 2021 Presented at: ICCOINS '21; July 13-15, 2021; Kuching, Malaysia p. 234-237 URL: https://ieeexplore.ieee.org/document/9497160 [doi: 10.1109/iccoins49721.2021.9497160]
59. Lokala U, Srivastava A, Dastidar TG, Chakraborty T, Akhtar MS, Panahiazar M, et al. A computational approach to understand mental health from Reddit: knowledge-aware multitask learning framework. Proc Int AAAI Conf Web Soc Media 2022 May 31;16:640-650. [doi: 10.1609/icwsm.v16i1.19322]
60. Ghoshal N, Bhartia V, Tripathy B, Tripathy A. Chatbot for mental health diagnosis using NLP and deep learning. In: Proceedings of the 2023 Conference on Advances in Distributed Computing and Machine Learning. 2023 Presented at: ICADCML '23; January 15-16, 2023; Rourkela, India p. 465-475 URL: https://link.springer.com/chapter/10.1007/978-981-99-1203-2_39 [doi: 10.1007/978-981-99-1203-2_39]

61. Iglesias M, Sinha C, Vempati R, Grace SE, Roy M, Chapman WC, et al. Evaluating a digital mental health intervention (Wysa) for workers' compensation claimants: pilot feasibility study. J Occup Environ Med 2023 Feb 01;65(2):e93-e99 [FREE Full text] [doi: 10.1097/JOM.0000000000002762] [Medline: 36459701]
62. Linden B, Tam-Seto L, Stuart H. Adherence of the #Here4U App - military version to criteria for the development of rigorous mental health apps. JMIR Form Res 2020 Jun 17;4(6):e18890 [FREE Full text] [doi: 10.2196/18890] [Medline: 32554374]
63. Shorey S, Ang E, Yap J, Ng ED, Lau ST, Chui CK. A virtual counseling application using artificial intelligence for communication skills training in nursing education: development study. J Med Internet Res 2019 Oct 29;21(10):e14658 [FREE Full text] [doi: 10.2196/14658] [Medline: 31663857]
64. Williams D, Hodge VJ, Wu CY. On the use of AI for generation of functional music to improve mental health. Front Artif Intell 2020 Nov 19;3:497864 [FREE Full text] [doi: 10.3389/frai.2020.497864] [Medline: 33733192]
65. Jameel U, Anwar A, Khan H. Doctor recommendation Chatbot: a research study. J Appl Artif Intell 2021 Dec 31;2(1):1-8. [doi: 10.48185/jaai.v2i1.310]
66. Liu YC, Xia S, Nie JP, Wei P, Shu Z, Chang JA, et al. aiMSE: toward an AI-based online mental status examination. IEEE Pervasive Comput 2022 Oct 1;21(4):46-54. [doi: 10.1109/mprv.2022.3172419]
67. Anastasiia M, Korotyeva T. A chatbot of a person's emotional state using a neural network. In: Proceedings of the 17th International Conference on Computer Sciences and Information Technologies. 2022 Presented at: CSIT '22; November 10-12, 2022; Lviv, Ukraine p. 13-16 URL: https://ieeexplore.ieee.org/document/10000558
68. Høiland CG, Følstad A, Karahasanovic A. Hi, can I help? Exploring how to design a mental health chatbot for youths. Hum Technol 2020 Aug 31;16(2):139-169. [doi: 10.17011/ht/urn.202008245640]
69. Nelekar S, Abdulrahman A, Gupta M, Richards D. Effectiveness of embodied conversational agents for managing academic stress at an Indian University (ARU) during COVID‐19. Brit J Educational Tech 2021 Dec 20;53(3):491-511. [doi: 10.1111/bjet.13174]
70. Adikari A, de Silva D, Moraliyage H, Alahakoon D, Wong J, Gancarz M, et al. Empathic conversational agents for real-time monitoring and co-facilitation of patient-centered healthcare. Future Gener Comput Syst 2022 Jan;126:318-329. [doi: 10.1016/j.future.2021.08.015]
71. Tran BX, McIntyre RS, Latkin CA, Phan HT, Vu GT, Nguyen HL, et al. The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis. Int J Environ Res Public Health 2019 Jun 18;16(12):2150 [FREE Full text] [doi: 10.3390/ijerph16122150] [Medline: 31216619]
72. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell 2023 Jan 23;5(1):46-57 [FREE Full text] [doi: 10.1038/S42256-022-00593-2]
73. Stapleton L, Taylor J, Fox S, Wu T, Zhu H. Seeing seeds beyond weeds: green teaming generative AI for beneficial uses. arXiv Preprint posted online May 30, 2023. [FREE Full text]
74. Farhat F. ChatGPT as a complementary mental health resource: a boon or a bane. Ann Biomed Eng 2024 May;52(5):1111-1114 [FREE Full text] [doi: 10.1007/s10439-023-03326-7] [Medline: 37477707]
75. Hou Y. AI music therapist: a study on generating specific therapeutic music based on deep generative adversarial network approach. In: Proceedings of the 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information. 2022 Presented at: ICETCI '22; May 27-29, 2022; Changchun, China p. 1277-1281 URL: https://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=9832398&ref= [doi: 10.1109/icetci55101.2022.9832398]
76. Perlis RH. Application of GPT-4 to select next-step antidepressant treatment in major depression. medRxiv Preprint posted online April 18, 2023 [FREE Full text] [doi: 10.1101/2023.04.14.23288595]
77. Wang Z. Generate Reflections and Paraphrases out of Distress Stories in Mental Health Forums. EPFL. 2022. URL: https://www.epfl.ch/labs/gr-pu/wp-content/uploads/2022/07/Generate-Reflections-and-Paraphrases-out-of-Distress-Stories-in-Mental-Health-Forums-1.pdf [accessed 2024-04-29]
78. Wei Z. Exploration of intelligent chatbots for auxiliary treatment of depression. Changsha University of Science and Technology. 2020. URL: https://chn.oversea.cnki.net/kcms/detail/detail.aspx?filename=1021046585.nh&dbcode=CMFD&dbname=CMFD202102&uniplatform=NZKPT [accessed 2024-04-29]
79. Tye MP. Multi-round dialogue model and application based on graph neural network. University of Electronic Science and Technology of China. 2021. URL: https://chn.oversea.cnki.net/kcms/detail/detail.aspx?filename=1021748784.nh&dbcode=CMFD&dbname=CMFD202201&uniplatform=NZKPT [accessed 2024-04-29]
80. Sun Y. Research and application of sequence-based intent recognition and response generation technology. East China Normal University. 2022. URL: https://chn.oversea.cnki.net/kcms/detail/detail.aspx?filename=1022500841.nh&dbcode=CMFD&dbname=CMFD202202&uniplatform=NZKPT [accessed 2024-04-29]
81. Zhuang Z. Psychological consultation chatbot based on GPT-2. Nanjing University of Posts and Telecommunications. 2022. URL: https://chn.oversea.cnki.net/kcms/detail/detail.aspx?filename=1022784732.nh&dbcode=CMFD&dbname=CMFD202301&uniplatform=NZKPT [accessed 2024-04-29]

82. Brocki L, Dyer GC, Gladka A, Chung NC. Deep learning mental health dialogue system. In: Proceedings of the 2023 IEEE International Conference on Big Data and Smart Computing. 2023 Presented at: BigComp '23; February 13-16, 2023; Jeju, Republic of Korea p. 395-398 URL: https://ieeexplore.ieee.org/document/10066740 [doi: 10.1109/bigcomp57234.2023.00097]
83. Burrows H, Zarrin J, Babu-Saheer L, Maktab-Dar-Oghaz M. Realtime emotional reflective user interface based on deep convolutional neural networks and generative adversarial networks. Electronics 2021 Dec 31;11(1):118. [doi: 10.3390/electronics11010118]
84. Elyoseph Z, Hadar-Shoval D, Asraf K, Lvovsky M. ChatGPT outperforms humans in emotional awareness evaluations. Front Psychol 2023;14:1199058 [FREE Full text] [doi: 10.3389/fpsyg.2023.1199058] [Medline: 37303897]
85. Lee D, Han SH, Oh KJ, Choi HJ. A temporal community contexts based funny joke generation. In: Proceedings of the 18th IEEE International Conference on Mobile Data Management. 2017 Presented at: MDM '17; May 29-June 1, 2017; Daejeon, South Korea p. 360-365 URL: https://ieeexplore.ieee.org/document/7962480 [doi: 10.1109/mdm.2017.62]
86. Li Y, Li XL, Lou Z, Chen CF. Long short-term memory-based music analysis system for music therapy. Front Psychol 2022 Jun 14;13:928048 [FREE Full text] [doi: 10.3389/fpsyg.2022.928048] [Medline: 35774954]
87. Liyanage C, Garg M, Mago V, Sohn S. Augmenting Reddit posts to determine wellness dimensions impacting mental health. Proc Conf Assoc Comput Linguist Meet 2023 Jul;2023:306-312 [FREE Full text] [doi: 10.18653/v1/2023.bionlp-1.27] [Medline: 38384674]
88. Rajagopal A, Nirmala V, Andrew J, Arun M. Novel AI to avert the mental health crisis in COVID-19: novel application of GPT2 in cognitive behaviour therapy. J Affect Disord 2021:1-42 [FREE Full text] [doi: 10.21203/rs.3.rs-382748/v3]
89. Rajcic N, McCormack J. Mirror ritual: an affective interface for emotional self-reflection. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 1-13 URL: https://dl.acm.org/doi/pdf/10.1145/3313831.3376625 [doi: 10.1145/3313831.3376625]
90. Ratican J, Hutson J. The six emotional dimension (6DE) model: a multidimensional approach to analyzing human emotions and unlocking the potential of emotionally intelligent artificial intelligence (AI) via large language models (LLM). DS J Artif Intell Robot 2023 Sep 30;1(1):44-51. [doi: 10.59232/air-v1i1p105]
91. Wang X, Guo B, Ma X, Bai X. Research on emotional music reconstruction method based on DBN-GRU. J Phys Conf Ser 2021 Jul 01;1966(1):012016. [doi: 10.1088/1742-6596/1966/1/012016]
92. Huang Z, Shen L. Algorithmic composition methods for wearable health devices. J Fudan Univ 2019;58(4):515-520 [FREE Full text]
93. Kumar H, Wang Y, Shi J, Musabirov I, Farb NA, Williams JJ. Exploring the use of large language models for improving the awareness of mindfulness. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Presented at: CHI EA '23; April 23-28, 2023; Hamburg, Germany p. 1-7 URL: https://dl.acm.org/doi/pdf/10.1145/3544549.3585614 [doi: 10.1145/3544549.3585614]
94. Lai T, Shi Y, Du Z, Wu J, Fu K, Dou Y, et al. Psy-LLMcaling up global mental health psychological services with AI-based large language models. arXiv Preprint posted online July 22, 2013. [FREE Full text]
95. Qiu H, He H, Zhang S, Li A, Lan Z. SMILE: single-turn to multi-turn inclusive language expansion via ChatGPT for mental health support. arXiv Preprint posted online April 30, 2023. [FREE Full text]
96. Roy K, Gaur M, Soltani M, Rawte V, Kalyan A, Sheth A. ProKnow: process knowledge for safety constrained and explainable question generation for mental health diagnostic assistance. Front Big Data 2022;5:1056728 [FREE Full text] [doi: 10.3389/fdata.2022.1056728] [Medline: 36700134]
97. Saha T, Chopra S, Saha S, Bhattacharyya P, Kumar P. A large-scale dataset for motivational dialogue system: an application of natural language generation to mental health. In: Proceedings of the 2021 International Joint Conference on Neural Networks. 2021 Presented at: IJCNN '21; July 18-22, 2021; Shenzhen, China p. 1-8 URL: https://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=9533924&ref= [doi: 10.1109/ijcnn52387.2021.9533924]
98. Saha T, Gakhreja V, Das AS, Chakraborty S, Saha S. Towards motivational and empathetic response generation in online mental health support. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022 Presented at: SIGIR '22; July 11-15, 2022; Madrid, Spain p. 2650-2656 URL: https://dl.acm.org/doi/10.1145/3477495.3531912 [doi: 10.1145/3477495.3531912]
99. Sharma A, Rushton K, Lin IW, Wadden D, Lucas KG, Miner AS, et al. Cognitive reframing of negative thoughts through human-language model interaction. arXiv Preprint posted online May 4, 2023 [FREE Full text] [doi: 10.18653/v1/2023.acl-long.555]
100. Tlachac M, Gerych W, Agrawal K, Litterer B, Jurovich N, Thatigotla S, et al. Text generation to aid depression detection: a comparative study of conditional sequence generative adversarial networks. In: Proceedings of the 2022 IEEE International Conference on Big Data. 2022 Presented at: BigData '22; December 17-20, 2022; Osaka, Japan p. 2804-2813 URL: https://ieeexplore.ieee.org/abstract/document/10020224 [doi: 10.1109/bigdata55660.2022.10020224]
101. Xu X, Yao B, Dong Y. Mental-LLM: leveraging large language models for mental health prediction via online text data. arXiv Preprint posted online July 26, 2023. [FREE Full text]
102. Yang K, Ji S, Zhang T, Xie Q, Ananiadou S. Towards interpretable mental health analysis with large language models. arXiv Preprint posted online October 11, 2023..

103. Zeberga K, Attique M, Shah B, Ali F, Jembre Y, Chung TS. A novel text mining approach for mental health prediction using Bi-LSTM and BERT model. Comput Intell Neurosci 2022;2022:7893775 [FREE Full text] [doi: 10.1155/2022/7893775] [Medline: 35281185]
104. Li Z, Chen Y, Zhang B, Chen L, Guo B. Stress-reducing music reconstruction based on deep learning. Comput Appl Softw 2021;39(8):158-162 [FREE Full text] [doi: 10.3969/j.issn.1000-386x.2022.08.024]
105. Singh OP. Artificial intelligence in the era of ChatGPT - opportunities and challenges in mental health care. Indian J Psychiatry 2023 Mar;65(3):297-298 [FREE Full text] [doi: 10.4103/indianjpsychiatry.indianjpsychiatry_112_23] [Medline: 37204980]
106. Teterwak P, Sarna A, Krishnan D, Maschinot A, Belanger D, Liu C, et al. Boundless: generative adversarial networks for image extension. arXiv Preprint posted online August 19, 2019. [FREE Full text] [doi: 10.1109/iccv.2019.01062]
107. Daws R. Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves. AI News. URL: https://www.artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/ [accessed 2024-04-29]
108. Timmons AC, Duong JB, Simo Fiallo N, Lee T, Vo HP, Ahle MW, et al. A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. Perspect Psychol Sci 2023 Sep 09;18(5):1062-1096 [FREE Full text] [doi: 10.1177/17456916221134490] [Medline: 36490369]
109. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. Proc Mach Learn Res 2018;8:77-91 [FREE Full text]
110. Brougham D, Haar J. Smart technology, artificial intelligence, robotics, and algorithms (STARA): employees' perceptions of our future workplace. J Manage Organ 2017 Jan 24;24(2):239-257. [doi: 10.1017/jmo.2016.55]
111. Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen L, et al. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health 2022 Jun;4(6):e406-e414. [doi: 10.1016/s2589-7500(22)00063-2]
112. Poon AI, Sung JJ. Opening the black box of AI-medicine. J Gastroenterol Hepatol 2021 Mar 11;36(3):581-584. [doi: 10.1111/jgh.15384] [Medline: 33709609]
113. Fui-Hoon Nah F, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration. J Inf Technol Case Appl Res 2023 Jul 21;25(3):277-304. [doi: 10.1080/15228053.2023.2233814]
114. Pan D, Sze S, Minhas JS, Bangash MN, Pareek N, Divall P, et al. The impact of ethnicity on clinical outcomes in COVID-19: a systematic review. EClinicalMedicine 2020 Jun;23:100404 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100404] [Medline: 32632416]
115. Jiao W, Chang A, Ho M, Lu Q, Liu MT, Schulz PJ. Predicting and empowering health for generation z by comparing health information seeking and digital health literacy: cross-sectional questionnaire study. J Med Internet Res 2023 Oct 30;25:e47595 [FREE Full text] [doi: 10.2196/47595] [Medline: 37902832]
116. Chang A, Schulz PJ, Jiao W, Liu MT. Obesity-related communication in digital Chinese news from Mainland China, Hong Kong, and Taiwan: automated content analysis. JMIR Public Health Surveill 2021 Nov 23;7(11):e26660 [FREE Full text] [doi: 10.2196/26660] [Medline: 34817383]


## Abbreviations

**AI:** artificial intelligence
**GAI:** generative artificial intelligence
**GPT:** generative pretrained transformer
**LSTM:** long short-term memory
**ML:** machine learning
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews


*Edited by T de Azevedo Cardoso; submitted 15.10.23; peer-reviewed by A Rovetta, T Ward; comments to author 19.01.24; revised version received 02.04.24; accepted 26.04.24; published 12.08.24.*

*Please cite as:*
*Xian X, Chang A, Xiang YT, Liu MT*
*Debate and Dilemmas Regarding Generative AI in Mental Health Care: Scoping Review*
*Interact J Med Res 2024;13:e53672*
*URL: https://www.i-jmr.org/2024/1/e53672*
*doi:10.2196/53672*
*PMID:*

Review

# Exploring Computational Techniques in Preprocessing Neonatal Physiological Signals for Detecting Adverse Outcomes: Scoping Review

Jessica Rahman[1], BSc, PhD; Aida Brankovic[2], PhD; Mark Tracy[3], PhD; Sankalp Khanna[2], PhD

[1]Commonwealth Scientific and Industrial Research Organisation (CSIRO) Australian e-Health Research Centre, Australia, Sydney, Australia
[2]Commonwealth Scientific and Industrial Research Organisation (CSIRO) Australian e-Health Research Centre, Australia, Brisbane, Australia
[3]Neonatal Intensive Care Unit, Westmead, Sydney, Australia

**Corresponding Author:**
Jessica Rahman, BSc, PhD
Commonwealth Scientific and Industrial Research Organisation (CSIRO) Australian e-Health Research Centre, Australia
160 Hawkesbury Road
Sydney, 2145
Australia
Phone: 61 02 9325 3016
Email: jessica.rahman@csiro.au

## Abstract

**Background:** Computational signal preprocessing is a prerequisite for developing data-driven predictive models for clinical decision support. Thus, identifying the best practices that adhere to clinical principles is critical to ensure transparency and reproducibility to drive clinical adoption. It further fosters reproducible, ethical, and reliable conduct of studies. This procedure is also crucial for setting up a software quality management system to ensure regulatory compliance in developing software as a medical device aimed at early preclinical detection of clinical deterioration.

**Objective:** This scoping review focuses on the neonatal intensive care unit setting and summarizes the state-of-the-art computational methods used for preprocessing neonatal clinical physiological signals; these signals are used for the development of machine learning models to predict the risk of adverse outcomes.

**Methods:** Five databases (PubMed, Web of Science, Scopus, IEEE, and ACM Digital Library) were searched using a combination of keywords and MeSH (Medical Subject Headings) terms. A total of 3585 papers from 2013 to January 2023 were identified based on the defined search terms and inclusion criteria. After removing duplicates, 2994 (83.51%) papers were screened by title and abstract, and 81 (0.03%) were selected for full-text review. Of these, 52 (64%) were eligible for inclusion in the detailed analysis.

**Results:** Of the 52 articles reviewed, 24 (46%) studies focused on diagnostic models, while the remainder (n=28, 54%) focused on prognostic models. The analysis conducted in these studies involved various physiological signals, with electrocardiograms being the most prevalent. Different programming languages were used, with MATLAB and Python being notable. The monitoring and capturing of physiological data used diverse systems, impacting data quality and introducing study heterogeneity. Outcomes of interest included sepsis, apnea, bradycardia, mortality, necrotizing enterocolitis, and hypoxic-ischemic encephalopathy, with some studies analyzing combinations of adverse outcomes. We found a partial or complete lack of transparency in reporting the setting and the methods used for signal preprocessing. This includes reporting methods to handle missing data, segment size for considered analysis, and details regarding the modification of the state-of-the-art methods for physiological signal processing to align with the clinical principles for neonates. Only 7 (13%) of the 52 reviewed studies reported all the recommended preprocessing steps, which could have impacts on the downstream analysis.

**Conclusions:** The review found heterogeneity in the techniques used and inconsistent reporting of parameters and procedures used for preprocessing neonatal physiological signals, which is necessary to confirm adherence to clinical and software quality management system practices, usefulness, and choice of best practices. Enhancing transparency in reporting and standardizing procedures will boost study interpretation and reproducibility and expedite clinical adoption, instilling confidence in the research findings and streamlining the translation of research outcomes into clinical practice, ultimately contributing to the advancement of neonatal care and patient outcomes.

*(Interact J Med Res 2024;13:e46946)* doi:10.2196/46946

**KEYWORDS**

physiological signals; preterm; neonatal intensive care unit; morbidity; signal processing; signal analysis; adverse outcomes; predictive and diagnostic models

# Introduction

## Background

Premature infants are those born at <37 weeks gestational age, ranging from extreme preterm (23 weeks' gestation) to late preterm (37 weeks' gestation), and are defined as having very low birth weight of <1500 g. These extremely premature infants have a higher risk of death, and surviving infants are highly prone to physical, cognitive, and emotional impairment [1]. The patients usually have a long length of stay, ranging from <10 to >120 days [2], in the neonatal intensive care unit (NICU), where high-fidelity physiological changes are monitored to observe their health status and signs of deterioration. During this long length of stay, a large amount of data from infants are generated and not typically electronically aggregated for permanent storage [3]. With the advent of electronic health records, relevant patient information is easily available for advanced data analytics that can be used to improve health outcomes. The records contain demographics, etiology, pathology, medication, and physiology information. Physiological changes are regularly monitored in preterm infants, notably, electrocardiogram (ECG), oxygen saturation ($SpO_2$), heart rate (HR), respiratory rate, arterial blood pressure, electroencephalography (EEG), and temperature. Some advanced centers around the world have started linking the information derived from the electronic health records data with the continuously monitored physiological information for permanent storage, more frequently in lower resolution, which facilitates various data analytics [4-6]. Compared with intermittent assessment and review, continuous capturing and analysis of the physiological data from the standard bedside monitors allow for a better understanding of trends and have been shown to improve outcomes of infants in the NICU [5].

Clinical decision support systems (CDSSs) can integrate clinical and physiological information to provide automated support in patient care planning to facilitate the diagnostic process and therapy planning, generate critical alerts and reminders, and predict the risk of patient deterioration. CDSSs have the potential for a positive impact in improving clinical and economic measures in the health care system [7-9]. The technological advancement that allowed storing big data, as well as the advancement of artificial intelligence (AI), has given rise to machine learning (ML)– and AI-based CDSSs aiming to build data-driven models to predict adverse outcomes in premature infants ahead of clinical diagnosis time [10-12].

The steps of building the ML pipeline to predict adverse outcomes involve several intermediate computational steps using the physiological data, of which data preprocessing is the first indispensable step. Namely, in the NICU, physiological signals are collected using a diverse range of devices, which introduce a number of artifacts such as *environmental artifacts* (eg, device connection failure, equipment noise, electrosurgical noise, and power line interferences); *experimental or human error* due to patient movement during data acquisition, incorrect or poor contact of the electrodes, and other contact noise; and *artifacts* due to muscle contraction, cardiac signals, and blinking [13,14]. These noises distort signals and may adversely affect model generalization capability and predictive power [10].

Although recently much progress has been made in building ML models using neonatal physiological data, there are limitations in the detailed reporting of the preprocessing techniques of these signals [15], which in turn hinder the reproducibility of the methods and results. In AI-powered software as a medical device (SaMD), this is especially important as the implementation of a software quality management system (QMS) is only possible by following the best practices and adhering to relevant regulatory standards and guidelines for medical devices, such as ISO 13485, IEC 62304, and IEC 82304-1. Beyond market access considerations, the ongoing international discourse on the regulation of medical software is specifically concentrated on AI and ML. This focus is a response to their growing applications, demanding increased attention from regulatory bodies such as the Australian Therapeutic Goods Administration and the US Food and Drug Administration [16]. Thus, it is crucial to adhere to a standardized protocol following clinical principles guided by domain experts and regulatory requirements while preprocessing the signals and reporting these techniques in detail; this ensures the reproducibility of the methods, allowing transparency in their clinical adoption.

## Objectives

As the first step in bridging the gap in their reproducibility for clinical adoption, this review aims to identify studies that used computational methods to analyze premature infants' physiological signals for detecting adverse outcomes. The review describes different tools and techniques used to preprocess physiological signals and provides recommendations on what aspects need further details for the clinical adoption of the techniques. The remainder of the paper is organized as follows: the Methods section explains the detailed search and screening process, while the Results section begins with an overview of the reviewed studies, followed by a detailed analysis. The Discussion section highlights the key reporting patterns identified in this review along with their shortcomings and provides recommendations for transparent reporting of future studies as it allows for accurate reproduction of the results and makes them usable in the clinical setting [17]. A summary of the work concludes the paper.

## Methods

### Search Strategy

The database searches and study screening were conducted following the recommendations of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [18] and the Centre for Reviews and Dissemination guidance for undertaking reviews in health care [19].

### Database and Search Strategies

A systematic database search was conducted on 5 databases: PubMed, IEEE, Web of Science, Scopus, and ACM Digital Library. The keywords were categorized into four concepts, which were then merged using the "AND" operator: concept 1—neonates or preterm infants; concept 2—vital signs or physiological signals; concept 3—computational techniques or signal processing; and concept 4—outcomes relating to neonates. Within each of these concepts, a combination of keywords and MeSH (Medical Subject Headings) terms were used to conduct the search process. The keywords under each concept were combined by the "OR" operator. The searches were limited to only the titles and abstracts. Table 1 shows the list of keywords and Medical Subject Headings terms used to search the database.

The search was done on January 9, 2023, and the publication year of the papers was limited to 2013 to 2023. The reason for choosing the 10-year range was to report on recent techniques and tools, as the devices and computational tools used >10 years ago may be obsolete. Scopus, Wiley Online Library, and Web of Science have an additional filter for choosing the subject area. This was used to restrict the subject areas to multidisciplinary, engineering, computing, and statistics. This was done to identify more papers on multidisciplinary areas through these databases, as PubMed covers all the major medical and health informatics databases. The combination of the 5 databases ensured that all medical, information technology, and multidisciplinary research papers were included in the database search. The search was restricted to English-language articles. Finally, review articles were excluded from the search.

**Table 1.** List of keywords and MeSH (Medical Subject Headings) terms used to conduct the database search.

| Concepts | Search strategy |
|---|---|
| **Concept 1: neonates or preterm babies** | |
| MeSH terms | "Infant, Premature" |
| Keywords | "premature" OR "preterm" OR "neonat*" OR "newborn" OR "infant" OR "nicu" OR "neonatal intensive care unit" |
| **Concept 2: physiological signals or vital signs** | |
| MeSH terms | "Vital Signs" OR "Physiology" |
| Keywords | "physiolog*" OR "ecg" OR "heart rate" OR "electrocardiography" OR "vital sign*" OR "physiomarker" OR "biomarker" OR "hrv" |
| **Concept 3: computational techniques or signal processing** | |
| MeSH terms | "Signal Processing, Computer-Assisted" |
| Keywords | "signal *" OR "predict*" OR "detect*" OR "comput*" |
| **Concept 4: outcomes** | |
| MeSH terms | None |
| Keywords | "sepsis" OR "mortality" OR "length of stay" OR "intraventricular hemorrhage" OR "hypoxi*" OR "apnea" OR "necrotising entercolitis" OR "necrotizing entercolitis" |

### Screening and Study Selection

The initial screening of the databases led to 3585 papers. Of these, 590 (16.46%) papers were manually identified as duplicates and excluded from the analysis. One paper was identified as a duplicate by the automation tool and removed. The remaining 2994 (83.51%) papers were subjected to title and abstract screening using the Rayyan Intelligent Systematic Review application (Qatar Computing Research Institute) [20].

Several inclusion criteria were set to select papers for full-text review. The criteria are mentioned in Textbox 1.

After screening the titles and abstracts, 81 articles were selected for full-text review; 29 (36%) papers were excluded during this stage as they did not align with the inclusion criteria, leaving 52 (64%) papers eligible for detailed synthesis and analysis.

The title and abstract screening was done by 1 reviewer, while 2 reviewers independently checked for paper eligibility against the inclusion criteria at the full-text review stage. When both reviewers were not in agreement on any papers, a third reviewer assessed them to provide a final decision on the inclusion and exclusion of the papers. Data charting was done using Microsoft Excel, and the following variables were recorded in line with related review papers [10,21]: title, year, journal, authors, digital object identifier, data set, participant number, participant demographic, signals used, data set size, sample rate, other data (if applicable), outcome metric, device software, programming language, preprocessing methods, algorithms, other techniques, features, models, model type, results (quantified), and key findings. Data synthesis was done using a narrative approach by summarizing findings based on the similarities in the data sets and techniques used. The detailed search queries,

bibliography files of all databases, all included papers, metadata of all papers and metadata of all papers included for full-text review are provided in Multimedia Appendices 1-5 [22-73].

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

**Inclusion criteria**

- Article type: articles must be peer-reviewed publications in a journal, conference, or workshop
- Data: articles must conduct an analysis on premature human infant data; articles must use physiological responses in some form
- Outcome: articles discuss applications relating to adverse neonatal outcomes such as mortality, length of stay, sepsis, necrotizing enterocolitis, intraventricular hemorrhage, hypoxic-ischemic encephalopathy, apnea, bradycardia, and other poor health outcomes, also known as morbidity. The disease outcomes were chosen based on the commonly researched outcome metric using preterm infant data and the search terms used in McAdams et al [10] that investigated artificial intelligence and machine learning techniques used to predict clinical outcomes in the neonatal intensive care unit
- Analysis: articles reported some form of computational techniques in their analysis
- Language: English

**Exclusion criteria**

- Article type: review papers are excluded
- Data: nonhuman data (eg, piglet infant data would not be considered); videos and images that do not look at the physiological responses and articles solely using demographic data for analysis were excluded
- Outcome: articles not focusing on these specified neonatal adverse outcomes were excluded
- Analysis: articles that only reported responses in their raw format were excluded
- Language: any languages other than English

# Results

## Overview of the Included Studies

Figure 1 shows the full process of database search and study selection using a PRISMA flow diagram.

Of the 52 selected articles, 24 (46%) studies focused on diagnostic models, while the rest (n=28, 54%) focused on prognostic models. These included journal articles (n=34, 65%), conference articles (n=17, 33%) and a workshop article (n=1, 2%). The most prominent physiological signals analyzed were ECG (n=36, 69%), $SpO_2$ (n=21, 40%), HR (n=16, 31%), respiration (n=16, 31%), BP (n=6, 12%), EEG (n=4, 8%), and temperature (n=3, 6%). While 8 (15%) studies used a combination of programming languages; others used MATLAB (n=6, 12%), Python (n=6, 12%), and R software (n=1, 2%), while the remaining studies (n=31, 60%) did not report what language was used. Physiological data monitoring and capturing was done using a range of systems, which subsequently impacted the sampling rate and quality of the data, thus leading to heterogeneity of the studies. The most commonly used devices for data capturing were Phillips Intellivue MP20, MP70, MP450, and MX800 machines [74] (n=14, 27%). Some other notable devices and software were BedMaster Ex System [75], NicoletOne EEG system [76], ixTrend, Phillips Data Warehouse connect [77], and Vuelogger patient monitoring system. The most commonly analyzed outcomes of interest were sepsis (n=20, 38%), apnea (n=17, 33%), bradycardia (n=13, 25%), mortality (n=7, 13%), and hypoxic-ischemic encephalopathy (n=5, 10%). It should be noted that 14 (27%) of the reviewed studies analyzed a combination of adverse outcomes.

As the studies were found to be heterogeneous in their study design and analysis techniques, a narrative approach was taken to summarize the studies and their key findings. The studies were grouped according to the homogeneity in terms of the data sets used and sorted by the publication year. This approach was inspired by the review article by Mann et al [78].

One of the noticeable patterns identified through the results reported in Table 2 is that the groups publishing studies using the same data set followed similar preprocessing techniques, although not at every step. For instance, studies using the ECG data from Cork University Maternity Hospital all used the same algorithm for QRS complex detection. However, they were diverse in their selection of filtering techniques and segmentation duration. Furthermore, they systematically failed to report detailed parameter settings for the QRS complex detection. While the approach of using similar preprocessing techniques helps maintain consistency to some extent, they do not confirm adhering to clinical practices identified from domain expert knowledge.

The QRS complex characteristics and RR intervals for neonates are different from those of adults and as such require an appropriate adjustment for QRS detection algorithms. This is a necessary first step for HR variability (HRV) analysis in neonates. However, a review published on neonatal HRV by Latremouille et al [15] revealed that given a lack of clear guidelines on neonatal vital signs and HRV analysis, several studies followed HRV analysis guidelines for adults published by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology [79]. Our review found that 16 (44%) out of the 36 studies analyzing ECG signals used the Pan-Tompkins algorithms for QRS complex detection. The original implementation of the

algorithm was based on the ECG characteristics of the adult population and therefore was preprocessed accordingly. Only 4 (25%) of those 16 studies reported adjustment of the original algorithm to adapt to neonates, of which only 2 provided specific modification details. In the absence of detailed reporting on the parameter settings, it is difficult to determine whether the settings adhered to neonatal waveform morphology. Incomplete reporting and lack of transparency hinder the understanding of the strengths and weaknesses of a study and limit its reproducibility and usability. Moreover, transparent and detailed reporting is required to confirm the adherence to regulatory compliance and is crucial for the clinical adoption of these methods.

Similar to the QRS complex in ECG signals, the acceptable ranges of physiological signals for neonates are also different from those of the adult population. This review found that no studies reviewed the acceptable ranges of the analyzed signals against any published guidelines, which could pose several limitations in the clinical adoption of the methods. This is consistent with another review looking into physiological vital sign ranges from 34 weeks gestational age, and it identified that several studies reported the means of vital signs instead of ranges, which makes the interpretation into clinical practice difficult [80]. Here, we recommend clear reporting and the use of physiological signal ranges that are clinically validated through published studies and textbooks [81-83].

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the database search and study selection.

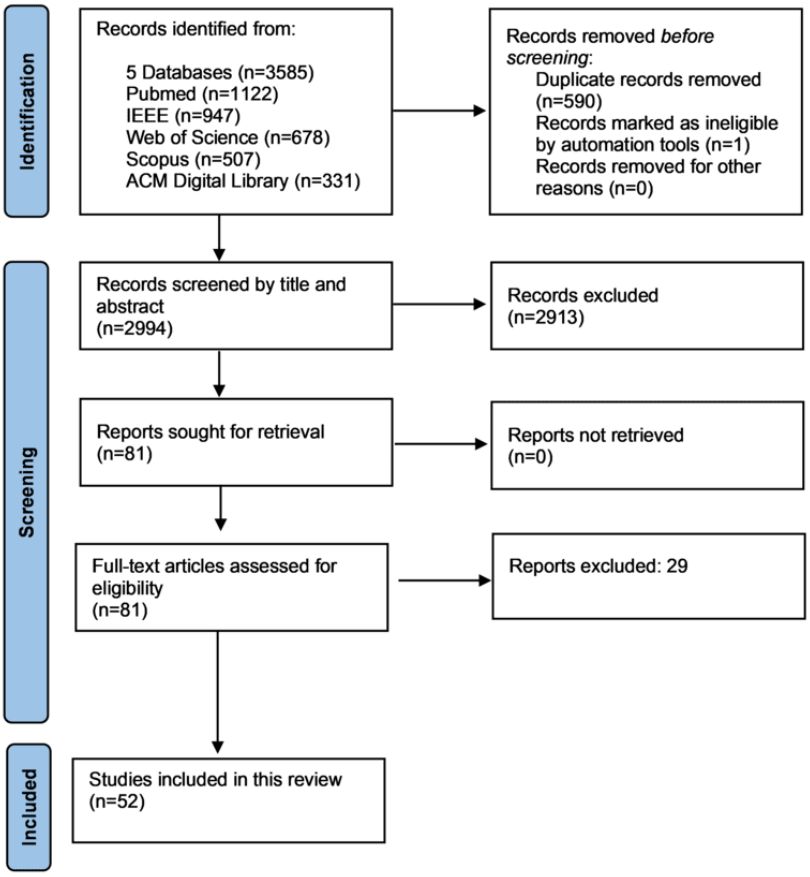


**Table 2.** Summary of the articles reviewed in this study, grouped according to the homogeneity in terms of the data sets used and sorted by the publication year.

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| National CHIME[a] database [84] | Cohen and de Chazal [22], 2013 | Participants: n=288; data size: NR[b]; model: diagnostic; outcome metric: sleep apnea | ECG[c] from single channel at 100 Hz, $SpO_2$[d] at 1 Hz | $SpO_2$ values <65% and changes in saturation exceeding 4% per second were discarded. ECG QRS complex was detected using the Pan-Tompkins algorithm [85] to generate RR intervals. QRS complexes were filtered using a technique from Chazal et al [86]. Filtered intervals were time aligned with $SpO_2$ using 30-second epochs | Eleven features were extracted from the signals. A combination of features from both signals resulted in 88.8% accuracy, 94.3% specificity, and 73.4% sensitivity in detecting sleep apnea |
| CHIME | Cohen and de Chazal [23], 2014 | Participants; n=402; data size: NR; model: diagnostic; outcome metric: sleep apnea | ECG from single channel at 100 Hz, $SpO_2$ at 1 Hz, actigraphy signals at 50 Hz | Actigraphy signals artifact rejection was done using the technique described by Lewicke et al [87]. $SpO_2$ values <65% and changes in saturation exceeding 4% per second were discarded. ECG data were passed through a QRS detection algorithm (NR) to produce RR intervals, which were filtered using a previously outlined method [86] | Fourteen features were extracted from the signals. A linear discriminant classifier achieved an accuracy of 74.1%, a sensitivity of 82.0%, and a specificity of 60.9% in detecting sleep apnea |
| CHIME | Cohen and de Chazal [24], 2015 | Participants; n=394; data size: NR; model: diagnostic; outcome metric: sleep apnea | ECG from single channel at 100 Hz, $SpO_2$ at 1 Hz | $SpO_2$ and ECG signals were time aligned to 30-second epochs. $SpO_2$ values <65% and changes in saturation exceeding 4% per second were discarded. ECG QRS complex detected using the Pan-Tompkins algorithm [85] to generate RR intervals. QRS complexes were filtered using a technique from Chazal et al [86] | Eleven features were extracted from both signals. A linear discriminant model achieved 66.7% accuracy, 67% specificity, and 58.1% sensitivity using features from both signals |
| PICS[e] database [25,88] | Gee et al [26], 2016 | Participants; n=10; data size: ~20-70 hours each; model: diagnostic; outcome metric: bradycardia | 3-lead ECG at 500 Hz, respiration signal at 50 Hz | RR intervals from ECG were extracted using a modified Pan-Tompkins algorithm (modification details NR). Analysis was done on a 3-minute window before each bradycardia. No processing was reported for respiration signals | Bradycardia severity estimation accuracy was improved by an average of 11% using a point process model of heart rate and respiration |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| PICS | Gee et al [25], 2017 | Participants, 10; data size: ~20-70 hours each; model: prognostic (+116 seconds); outcome metric: bradycardia | 3-lead ECG at 500 Hz, respiration signal at 50 Hz | RR intervals from ECG were extracted using a modified Pan-Tompkins algorithm (modification details NR). The artifacts, due to movement, disconnection, or erroneous peaks, were removed by visual inspection. No processing was reported for respiration signals. Additional analysis on the frequency content of the RR time series was done using Morlet wavelet transform [89] | A point process model–based prediction algorithm achieved a mean AUROC[f] of 0.79 for >440 bradycardic events and was able to predict bradycardic events on an average of 116 seconds before onset (FPR[g]=0.15) |
| PICS | Das et al [27], 2019 | Participants; n=10; data size: ~20-70 hours each; model: prognostic (time NR); outcome metric: bradycardia | 3-lead ECG at 500 Hz | Baseline wander was removed using a high-pass filter with a cutoff frequency between 0.5 and 0.6 Hz. Motion and disconnection artifacts were removed by visual inspection. QRS complexes were detected using Pan-Tompkins algorithm [85]. Signals were segmented 5 minutes before and 2 minutes after a bradycardic event | Nonparametric modeling using kernel density estimation achieved a 5% false alarm rate in predicting the onset of bradycardia events |
| PICS | Mahmud et al [28], 2019 | Participants; n=11; data size: ~20-70 hours each for 10 and 10 weeks for 1 participant; model: prognostic (time NR); outcome metric: bradycardia | 3-lead ECG at 500 Hz | QRS complex was detected using an algorithm (NR). RR intervals were calculated from the detected peaks | Time and frequency domain features were extracted. An extreme gradient boosting model achieved an average AUROC of 0.867. HRV[h] results showed a significant variation between a healthy infant and an infant prone to bradycardia |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| PICS | Gee et al [29], 2019 | Participant, n=10; data size: ~20-70 hours each; model: diagnostic; outcome metric: bradycardia, apnea of prematurity | 3-lead ECG at 500 Hz, respiration signal at 50 Hz | Respiration signals were clipped into 60-second segments and normalized to 0-mean, unit variance. RR intervals from ECG signals were extracted using a Morlet wavelet transformation. An open-source peak finder (name NR) was applied to the wavelet scale ranging from 0.01 to 0.04, which is related to the QRS complex formation in the spectrogram. ECG signals were segmented to 15 seconds with the event in the middle. The segments were bandpassed filtered from 3 to 45 Hz, scaled to 0-mean unit variance, and scaled to the median QRS complex amplitude. Waveforms were visually inspected to remove segments with no distinguishable QRS complex or respiratory peaks | An autoencoder-prototype model was proposed, which achieves 93.1% (SD 0.4%) accuracy in predicting bradycardia and 82.3% (SD 3.8%) accuracy in classifying apnea |
| MIMIC-III[i] database from Beth Israel Deaconess Medical Center [90] | Song et al [30], 2020 | Participants: 2819 (21 sepsis, 2798 control); data size: NR; model: prognostic (+48 hours); outcome metric: sepsis | HR[j], SBP[k], DBP[l], MBP[m], $SpO_2$, respiration, temperature, other (sampling rate NR) | Data quality was assessed by missing value filter and 3-sigma rule. The final observation carried forward was applied to vital signs not meeting data quality. Zero imputation was performed if calculation could not be performed (eg, divided by 0) | Several statistical features were extracted at 3-, 6-, 12-, and 24-hour window. Linear model, naive Bayes, decision tree, ensemble method, and neural network models were evaluated. The AUROC of the 48-hour prediction model achieved 0.861 and that of the onset detection model was 0.868 |
| MIMIC-III | Baker et al [31], 2021 | Participants; n=179 for 3-day and n=181 for 14-day model; data size, NR; model: prognostic (+3 days); outcome metric: mortality | HR, respiration signal, sampled hourly | Values <0 and flatline cases were eliminated | Several statistical features were extracted from the signals. CNN-LSTM[n] model using a 3-day scheme achieved AUROC of 0.9336 (SD 0.0337) across 5-fold cross-validation |
| MIMIC-III | Juraev et al [32], 2022 | Participants; n=3133; data size: 24 hours from each; model: prognostic (time NR); outcome metric: mortality and LOS[o] | HR, respiration signal, $SpO_2$, BP, temperature (sampling rate NR) | Missing data were filled by forward and backward filling, using the mean value. For participants >24 measurements, they were reduced by taking the average of the nearest records. For <24 measurements, values were generated using filling algorithm | A dynamic ensemble KNN[p] method reached 0.988 (SD 0.001) $F_1$-score in mortality classification. Voting of static ensemble regression models achieved an RMSE[q] of 12.509 (SD 0.079) in LOS prediction |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| University Hospitals in France | Ghahjaverestan et al [33], 2015 | Participants; n=32; data size: 105 segments of ECG with 250 seconds duration; model: diagnostic; outcome metric: apnea- bradycardia | One lead ECG at 400 Hz and respiration signals (sampling rate NR) | Baseline and noise of 50 Hz were removed from ECG signals, QRS complexes were detected using Pan-Tompkins algorithm [85]. RR intervals were further downsampled to 10 Hz for 1 prediction model | A Kalman-filter–based method achieved sensitivity and specificity of 94.74% and 94.17%, respectively, in predicting apnea-bradycardia episodes |
| University Hospitals in France | Navarro et al [34], 2015 | Participants; n=51; data size: testing cohort mean duration—2.4 hours; model: diagnostic; outcome metric: sepsis | Respiration signals at 400 Hz, downsampled to 64 Hz | Frequency content >32 Hz from breathing signals was removed using a seventh-order Butterworth low-pass filter. After rejecting artifacts due to gross movements, a fourth-order Butterworth filter with a cutoff frequency between 0.5 and 20 Hz was applied. Smoothing filtering using an $SG^r$ filter [91] was applied. A simple extrema detector is then applied to detect respiratory cycles | 14 features, computed in 10-second sliding excerpts, were extracted from the breathing signals. A logistic regression classifier automatically rejects artifacts to 86% sensitivity and specificity, which is used in the proposed framework for neonatal sepsis detection |
| University Hospitals in France | Ghahjaverestan et al [35], 2016 | Participants; n=32; data size: real (236 segments Synthetic) 200 sequences of 400 seconds; model: diagnostic (0.59-second delay); outcome metric: apnea-bradycardia | ECG at 400 Hz. Synthetic signals at 10 Hz | Baseline and noise of 50 Hz were removed from ECG signals using a combination of low-pass and notch filters; QRS complexes were detected using Pan-Tompkins algorithm [85]. Three features were extracted using a wavelet-based beat delineator [92]. Features were transformed to 10 Hz using interpolation (technique NR) | A $CHMM^s$ achieved 95.74% sensitivity and 91.88% specificity in detecting apnea-bradycardia episodes, with a detection delay of –0.59 seconds |
| University Hospitals in France | León et al [36], 2021 | Participant, n=49; data size: NR; model: prognostic (+6 hours); outcome metric: sepsis | ECG at 500 Hz | RR intervals were detected using modified Pan-Tompkins algorithms, and filter coefficients were adapted for newborns [93]. A sliding window of 30 minutes, with no overlaps, was applied to extract HRV parameters from the RR time series. 30-minute segments with a maximum RR >1 second or a minimum RR of <0.19 seconds were excluded, | Time, frequency, and nonlinear features were extracted from the HRV parameters. A logistic regression model using visibility graph features achieved 0.877 AUROC in predicting sepsis 6 hours before the start of antibiotics |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| University Hospitals in France | León et al [37], 2021 | Participants; n=259; data size: NR; model: prognostic (+6 hours); outcome metric: sepsis | ECG at 500 Hz | RR intervals were detected using modified Pan-Tompkins algorithms and filter coefficients were adapted for neonates [93]. RR time series were extracted and segmented into 5-minute segments. The 5-minute periods corresponding to 30 continuous minutes were grouped by calculating the median of each corresponding HRV feature | Time, frequency, nonlinear, and visibility graph features were extracted from the HRV parameters. An RNN[t] model achieved 0.904 AUROC in predicting sepsis 6 hours before the time of infection and >80% accuracy 24 hours before the onset of infection |
| University Hospitals in France | Doyen et al [38], 2021 | Participants; n=52; data size: 8 hours of recording from each; model: diagnostic (+2.9-second delay); outcome metric: bradycardia | 3-lead ECG at 300 Hz | QRS complexes were detected using a multifeature probabilistic real-time detector [93] | A high rate of false alarms (64%) was observed in real life. The proposed optimal decentralized fusion of 3 detection methods had a significant detection delay of 2.9 seconds, sensitivity of 97.6% and false alarm rate of 63.7% |
| University Hospitals in France | Sadoughi et al [39], 2021 | Participants; n=32; data size: 233 episodes with a duration of 21.48 (SD 16.07) seconds; model: diagnostic (+5.05-second delay); outcome metric: apnea-bradycardia | One lead ECG at 400 Hz | The same preprocessing techniques as reported in Ghahjaverestan et al [35]. QRS complexes were identified using Pan-Tompkins method [85]. The RR time series were uniformly upsampled to 10 Hz using a linear interpolation technique | A proposed layered HMM[u] model achieved 97.14% (SD 0.31%) accuracy in detecting apnea-bradycardia episodes, with a detection delay of –5.04 (SD 0.41) seconds |
| Cork University Maternity Hospital | Ahmed et al [40], 2015 | Participants: NR; data size: 54 1-hour recordings; model: diagnostic; outcome metric: HIE | 2-lead ECG, EEG[v] (sampling rate NR) | Artifacts were manually removed. R-peaks from raw ECG signals were extracted using Pan-Tompkins method [85]. The timing of the peaks was adjusted and uniformly sampled to 256 Hz using Hermite spline quadratic interpolation. Then, HRV features were extracted from a 1-minute window with 30-second overlap using the normalized RR interval | Seven time and frequency domain HRV features were extracted. A Gaussian supervector approach with SVM[w] achieved 0.81 AUROC in classifying HIE[x] |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| Cork University Maternity Hospital | Temko et al [41], 2015 | Participant, n=38; data size: 1 hour of EEG and ECG recordings from each; model: diagnostic; outcome metric: HIE | ECG and video EEG at 256 Hz | The 1-hour EEG segments were downsampled to 32 Hz with an antialiasing filter set to 16 Hz. The filtered EEG was segmented into a 60-second epoch with no overlap. QRS complexes from ECG signals were extracted using the algorithm reported in [94]. The resulting peaks were manually inspected to correct ectopic beats or mark artifacts. Then, signals were segmented into 60-second epochs | An SVM classifier using a subset of 9 EEG, 2 hours, and 1 clinical feature achieved 87% AUROC and 84% accuracy in predicting HIE |
| Cork University Maternity Hospital | Lloyd et al [42], 2016 | Participant, n=43; data size: mean recording duration 41 hours 40 minutes; model: diagnostic; outcome metric: future adverse outcome in infants | EEG at 256 Hz, $SpO_2$ and HR at 1 Hz | EEG recordings were visually checked for quality, and poor-quality data were discarded. 1-hour epochs of EEG at 12 and 2 hours of age were then extracted from each recording. 1-hour epochs of HR and $SpO_2$ were extracted at 12- and 24-hour time point. | A logistic regression model predicted a 2-year poor outcome with an AUROC of 0.83 |
| Cork University Maternity Hospital | Semenova et al [43], 2018 | Participants; n=35 with 23 used; data size: 824 hours; prognostic (time NR); outcome metric: short-term adverse outcome | ECG at 256 or 1024 Hz, BP at 1 Hz | Diastolic and systolic pressures every second were used to calculate $MAP^y$. ECG signals were segmented to nonoverlapping 5-minute epochs. QRS complexes were extracted by the Pan-Tompkins method [85]. Abnormal RR intervals were corrected by moving average. Periods of clear movement of artifacts were automatically discarded (method NR) | Fifteen time, frequency, and nonlinear features were extracted from HRV. An XGBoost decision tree using all features achieved an AUROC of 0.97 in predicting short-term outcomes in infants |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| Cork University Maternity Hospital | Semenova et al [44], 2019 | Participants; n=43 with 23 used; data size: total 831 hours; prognostic (time NR); outcome metric: 5 adverse outcomes | ECG at 256 or 1024 Hz, BP at 1 Hz | DBP and SBP every second were used to calculate MAP. Segments with MAP<10 mm Hg were discarded due to disconnection of the pressure transducer or movements. The MAP was segmented into 1-hour windows. Values outside 3 SD were discarded. ECG signals were segmented into nonoverlapping 5-minute epochs. QRS complexes were extracted using the Pan-Tompkins method [85]. ECG signal was bandpass filtered with 4-30-Hz cutoff frequency. Abnormal values of RR intervals were corrected by the moving average filter | Time, frequency, and nonlinear features were extracted from HRV. An XGBoost decision tree using a single HRV feature achieved 0.87 AUROC, while multiple features reached 0.97 AUROC in predicting adverse outcomes |
| Máxima Medical Center NICU[z] | Joshi et al [45], 2020 | Participants; n=49; data size: ~144 hours each; model: prognostic (+0-24 h); outcome metric: sepsis | ECG at 250 Hz, CI[aa] at 62.5 Hz | Respiration waveforms were bandpass filtered between 0.45 and 1.45 Hz. QRS complexes from ECG were extracted using a DT-CWT[ab]–based method described in [95].<br>IBIs[ac] were detected from the CI signal peaks using an algorithm (NR). Features were extracted from every 3-hour data | Twenty-two features were extracted from the signals. A naive bayes classifier reached up to 0.78 AUROC and 3 hours leading up to sepsis |
| Máxima medical Center NICU | Varisco et al [46], 2021 | Participants; n=20; data size: ~570 hours; model: prognostic (+6 hours); outcome metric: central apnea preceding late-onset sepsis | ECG at 240 or 250 Hz, CI at 60 or 62.5 Hz, $SpO_2$ at 0.5 or 1 Hz | A filtered respiration signal without cardiac artifacts was generated using algorithms reported in studies by Lee et al [96], Mohr et al [97], and Vergales et al [98]. Steps include Fourier transformation and integer frequencies filtered out, then resampled to 60 Hz and high-pass filtered with a cutoff frequency of 0.4 Hz, and a low-pass filter with a very low cutoff frequency optimized to fit apnea annotations by clinical experts (value NR) | An optimization of the algorithm was proposed to detect central apnea, which achieved 90.5% recall, 19.7% precision, and 30.8% $F_1$-score |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| Máxima medical Center NICU | Cabrera-Quiros et al [47], 2021 | Participants; n=64; data size: NR; model: prognostic (+3 hours); outcome metric: sepsis | ECG at 250 Hz, CI at 62.5 Hz | QRS complexes from ECG were extracted using a DT-CWT–based method described the same as Joshi et al [45]. CI signal was filtered to remove cardiac artifacts, and peaks were detected using methods similar to those in previous works (NR). Features were extracted from every 1-hour signal | Time domain features were extracted from HRV. Classification using a combination of all features and logistic regression model reached a mean accuracy of 0.79 (SD 0.12) and mean precision of 0.82 (SD 0.18), 3 hours before the onset of sepsis |
| Máxima Medical Center NICU | Varisco et al [48], 2022 | Participants; n=20; data size: 960 hours of data from 20 infants, 7818 event extracted; model: diagnostic; outcome metric: central apnea | ECG at 250 Hz, CI at 62.5 Hz, $SpO_2$ at 1 Hz | QRS complexes were detected using the same method as reported in Joshi et al [45] and Cabrera-Quiros et al [47]. From ECG, $SII^{ad}$ was calculated by applying a bandpass filter (0.001-0.40 Hz) using 10-second segments and then computing a kernel density estimate to return patient motion measurement every second. RR intervals were resampled at 250 Hz. CI signal was processed using the method by Redmond et al [99] to calculate $RRE^{ae}$. No preprocessing was done on $SpO_2$. Each feature was extracted using 30-second windows. *z* score normalization was applied to the feature matrix | 47 features were extracted from the vitals. A logistic regression model achieved 0.9 AUROC in detecting central apnea |
| Máxima Medical Center NICU | Peng et al [49], 2022 | Participants; n=128; data size: ~24 hours each; model: prognostic (+24 hours); outcome metric: sepsis | ECG at 250 Hz | QRS complexes from ECG were extracted using a DT-CWT–based method described by Rooijakkers et al [95]. RR intervals from the complexes were divided into nonoverlapping 1-hour segments. The segments were centered, and missing values in the segments were filled by zero padding on the 2 ends | A ResNet-based neural network, DeepLOS, was proposed, which achieved a 0.72 $F_1$-score in predicting late-onset sepsis |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| Máxima Medical Center NICU | Peng et al [50], 2022 | Participants; n=127; data size: ~48 hours each; model: prognostic (+6 hours); outcome metric: sepsis | ECG at 250 Hz, CI at 62.5 Hz | QRS complexes from ECG were extracted using a DT-CWT–based method described in [95]. CI signal was filtered to remove cardiac artifacts (method NR). Peaks were detected using the method reported by Lee et al [96]. SII was calculated from ECG and CI waveforms using a CWT-based method, as reported by Zuzarte et al [100]. Signals were divided into 1-hour-long nonoverlapping segments. Features were calculated in both 1-hour segments and 5-minute subsegments | 60 Features were extracted from the signals. An XGB model using the features achieved an AUROC of 0.88 in predicting late-onset sepsis 6 hours preceding the onset. |
| Royal Infirmary of Edinburgh NICU | Stanculescu et al [51], 2014 | Participants; n=24; data size: 30 hours each; model: prognostic (+3-6 hours); outcome metric: sepsis | ECG-derived HR, PR[af] (sampling rate NR) | An extension of the forward-backward algorithm [101] is developed for missing data inference | An autoregressive HMM model achieved up to 0.80 AUROC in predicting sepsis |
| Royal Infirmary of Edinburgh NICU | Stanculescu et al [52], 2014 | Participants; n=24; data size: 540 hours; model: diagnostic; outcome metric: sepsis | ECG-derived HR, PR core and peripheral temperature and $SpO_2$ at 1 Hz | An automated oximeter error detection algorithm was applied on the basis of the method described by Stanculescu et al [51]. Rows containing missing data on the observation matrix are set to 0 | An HSLDS[ag] was able to predict sepsis with up to 0.65 $F_1$-score |
| Kasturba hospital NICU, Manipal, India | Shirwaikar et al [53], 2016 | Participant: NR; data size: 229 examples; model: diagnostic; outcome metric: apnea | HR (sampling rate NR) | Visualization technique was applied to identify issues in data. Missing values were not treated due to low percentage. For categorical features, 0 was added for missing values. Minimum-maximum normalization and *z* score normalization were done | An RF[ah] model using HR features achieved 0.88 accuracy and 0.72 κ in detecting apnea |
| Kasturba hospital NICU, Manipal, India | Shirwaikar et al [54], 2019 | Participants; n=367 (315 used); data size: NR; model: diagnostic; outcome metric: apnea | ECG (sampling rate NR) | No preprocessing techniques were reported on the raw signals. Observations with missing features were discarded. Other features (continuous values) that had missing values were converted to discrete with the addition of the group name “not known” | Statistical features were extracted from the signals. A Multilayer Perceptron model and a deep autoencoder model reached 0.82 and 0.83 AUROC, respectively, in detecting apnea |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| University of Massachusetts Memorial Healthcare NICU | Williamson et al [55], 2013 | Participants; n=6; data size: ~5-8 hours for each patient; model: prognostic (+5.5 minutes); outcome metric: apnea | ECG, $SpO_2$, respirator signal, pulse plethysmogram (sampling rate NR) | IBIs were extracted from abdominal respiratory movements (method NR), and RR intervals were extracted from ECG signals (method NR). Physically implausible IBI and RR interval values were automatically removed (range NR). Values were resampled to 10 Hz using shape-preserving piecewise cubic interpolation. Signals were then log transformed and converted to 0 mean, unit variance | Features were extracted from all signals. A GMM[ai] model reached 0.8 AUROC in predicting apnea |
| Jackson Memorial Hospital NICU | Schiavenato et al [56], 2013 | Participants; n=20; data size: 1186 minutes; model: diagnostic; outcome metric: periods of high distress or pain | ECG at 1000 Hz | Pan-Tompkins algorithm [85] was modified to detect QRS complexes. ECG was filtered using a bandpass filter with a 16-26-Hz cutoff frequency. A low-pass filter by an order 120 FIR[aj] filter with a corner frequency of 25 Hz and a high-pass filter by an order 160 FIR filter with a corner frequency of 25 Hz were applied. Then, a polynomial filter of order 21 was applied as the differentiator filter. Finally, a 111-order moving average filter was used, and QRS complex was detected using an adaptive threshold. Lomb-Scargle LMS[ak] spectral estimation [102] was used for missing and irregular RR intervals | The proposed framework provided real-time analysis and HRV extraction to identify the characteristics correlated to periods of high distress or pain |
| Montreal Children's Hospital | Rubles-Rubio et al [57], 2014 | Participants; n=24; data size: 9.0 (SD 2.2) hours for each; model: diagnostic; outcome metric: apnea | $SpO_2$, RIP[al] (sampling rate NR) | Signals were low-pass filtered with a cutoff frequency of 10 Hz, with an 8-pole Bessel antialiasing filter digitized and sampled at 50 Hz | A linear Gaussian discriminant classifier detected the episodes with a 0.73 probability of detection and 0.22 probability of false alarm |
| University of Alabama at Birmingham | Amperayani et al [58], 2017 | Participants; n=18; data size: 24 hours each; model: prognostic (+23 hours); outcome metric: bradycardia, hypoxemia | ECG at 500 Hz and HR at 1 Hz | HR data were converted to interbeat RR intervals using RR=60/HR. No processing on ECG signals was reported | A point process model using RR intervals showed a strong correlation with bradycardia events and a modest correlation with hypoxemia events |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| Monash Children's Hospital NICU, Australia | Hu et al [59], 2018 | Participants; n<80; data size: 407 patient-day; model: prognostic (+24 hours); outcome metric: sepsis | HR, $SpO_2$, respiration signal at 1 Hz | Data were scaled down to 1 record per minute. Data blocks with invalid values were deleted. Then, the sliding window was set to 60 minutes to feed to the ML[am] models | Features were extracted from all signals. A gradient boosting decision tree achieved up to 0.97 AUROC and 0.92 weighted $F_1$-score in patient-based cross-validation in predicting sepsis |
| University of Virginia and Columbia University NICU | Sullivan et al [60], 2018 | Participants; n=78; data size: NR; model: prognostic (+12 hours); outcome metric: death, sIVH[an] (severe), BPD[ao], treated ROP, [ap] late-onset sepsis, and NEC[aq] | HR, $SpO_2$ at 0.5 Hz | Infants with <6 hours of data within 12 hours of birth were discarded. Cross-correlation of HR and $SpO_2$ was calculated over 10-minute windows using the XCORR function of MATLAB with a lag time of –30 to +30 seconds | A POPS[ar] was developed and fit a multivariate logistic regression model, which performed well in predicting death, sIVH, and BPD, but not tROP, sepsis, and NEC |
| 9 NICUs in the United States | Zimmet et al [61], 2020 | Participants; n=2989; data size: 121 data points per infant; model: prognostic (+2 days); outcome metric: mortality, sepsis | HRC[as] index from ECG | Infants with missing data on either end of the total duration were extrapolated to the window edge by repeating the most proximal HRC index values. Interior missing values were updated using linear interpolation. A fifth-order B-spline with equally spaced knots was used to capture information from independent samples (HRC indexes 12 samples apart) | An unsupervised ensemble of clustering techniques was proposed to cluster infants to different levels of risk |
| Children's National Hospital, Washington | Kota et al [103], 2020 | Participants; n=95; data size: median recording duration of 75.78 hours; model: diagnostic; outcome metric: HIE | EEG at 200 or 256 Hz | ECG contamination from EEG was detected using the method described by Govindan et al [104]. EEG signals with amplitude>500 µV or SD<0.01 µV were discarded as artifacts. The volume conduction was attenuated by calculating the global average of EEG voltages from all electrodes and subtracting the global average from the EEG value of every electrode in the frequency domain [62]. The values were then transformed to the time domain for spectral analysis. EEG was segmented into 10-minute nonoverlapping artifact and seizure-free epochs. Spectral analysis was done using a Welch periodogram approach [105,106] using 3-second epochs | EEG delta power was identified to be a crucial biomarker for predicting neonates with HIE who died with those who survived |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| Akbar Abadi Hospital NICU, Iran | Mirnia et al [63], 2021 | Participants; n=5; data size: ~24 hours each; model: diagnostic, outcome metric: sepsis | ECG at 200 Hz | RR intervals were calculated from ECG using HeRO[at] model | Features were extracted from HRV. HeRO model was tested using this data set. HeRO score was able to distinguish between healthy and septic newborns |
| St Louis Children's Hospital NICU | Lee et al [64], 2021 | Participants; n=275; data size: 4, 01,33,460 data points; model: prognostic (+6 hours); outcome metric: mortality | HR, respiration signal and $SpO_2$ at 1 Hz | Missing or out-of-range values were replaced with NaN and then imputed using mean values for that variable across all training and testing data. Data were downsampled to every 10 seconds to extract features. Dynamic variables were calculated as rolling means, SD, and absolute *z* score on 5- and 30-minute windows to reduce the influence of outliers | Thirty-four features were extracted from the signals. An RF model achieved 88% sensitivity and 0.93 AUROC in predicting mortality |
| University of Virginia Children's Hospital, Morgan Stanley Children's Hospital, and St Louis Children's Hospital | Sullivan et al [65], 2021 | Participants; n=408, (266 used); data size: NR; model: diagnostic; outcome metric: sepsis | HR and $SpO_2$ at 0.5 Hz | HR and $SpO_2$ values of 0 were removed. Eight features were extracted in 10-minute windows and averaged hourly. Cross-correlation between HR and $SpO_2$ was calculated in 10-minute windows of signals normalized to have 0 mean and SD of 1. Cross-correlation was done using the XCORR function of MATLAB with a lag time of –30 to +30 seconds | A logistic regression model using clinical and physiological features achieved an AUC[au] of 0.821 in predicting late-onset sepsis |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| St Louis Children's Hospital NICU | Feng et al [66], 2021 | Participants; n=285; data size: ~80 hours each; model: prognostic (+6 hours); outcome metric: mortality | HR, respiration, $SpO_2$, and ART-$M^{av}$ or NIBP-$M^{aw}$ at 1 Hz | Infants with data >80 hours were truncated, and <80 hours were padded with 0s. Mean, median, mode, and Bayesian ridge data imputation techniques were explored. Bayesian ridge was used to sample 5 data sets by sampling different posteriors each time. Then, the average was reported using 4-fold cross-validation. The rolling mean of each vital sign with a range of 5 minutes was used to reduce noise. Finally, the end of each sample was padded with 1 segment where all features equaled 0. Features were extracted from 5-minute segments | A deep learning model using LSTM named DeepPBSMonitor was developed to predict mortality with 0.888 accuracy, 0.78 recall, and 0.897 AUC |
| University of Massachusetts Memorial Healthcare | Zuzarte et al [67], 2021 | Participants; n=10; data size: 241.34 hours; model: prognostic (+310 seconds); outcome metric: apnea-bradycardia-hypoxia | ECG at 500 Hz, $PPG^{ax}$ at 125 Hz, $SpO_2$, HR, respiration signals from pneumogram at 50 Hz | PPG signals were filtered using a wavelet-based algorithm to remove gross body movements. A binary marker sampled at 25 Hz was obtained to indicate the presence or absence of movement. QRS complexes were detected using a modified Pan-Tompkins algorithm (modification NR). IBIs were detected using automated peak detection from LabChart Software RR intervals, and IBI values were then interpolated at 10 Hz | The prediction framework using GMM and logistic regression model achieved 75% accuracy in predicting bradycardia severity during the apnea-bradycardia-hypoxia event |
| University of Virginia NICU | Niestroy et al [68], 2022 | Participants; n=5957; data size: random daily 10 minutes segments from each; model: prognostic (+1-7 days); outcome metric: mortality | HR and $SpO_2$ at 0.5 Hz | No preprocessing was reported on the vitals. They were grouped to calculate the average in 10-minute nonoverlapping windows | Features were extracted from all signals. A multivariable logistic regression model using 5 features achieved the AUROC of 0.83 in predicting mortality |
| University of Virginia Children's Hospital, Morgan Stanley Children's Hospital and St Louis Children's Hospital | Kausch et al [69], 2023 | Participants; n=2494; data size: NR; model: prognostic (+24 hours); outcome metric: sepsis | HR and $SpO_2$ at 0.5 Hz | HR and $SpO_2$ were preprocessed by removing the values containing 0. Features were calculated in 10-minute nonoverlapping windows. Windows with >50% missing data were excluded from subsequent analysis | Several features were extracted from the vitals. An XGB model achieved training AUROC of 0.834 using the data from NICU 1, and 0.792 and 0.807 testing AUROC using data from NICU 2 and NICU 3, respectively |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| Karolinska University Hospital Solna and Huddinge NICU, Stockholm, Sweden | Honoré et al [70], 2023 | Participants; n=325; data size: 2866 hospitalization days; model: prognostic (+24 hours); outcome metric: sepsis | IBI from ECG, respiration from CI, $SpO_2$ (sampling rate 1-500 Hz) | All signals were resampled to 1 Hz. Segments with at most 15 seconds missing were linearly interpolated. All signals were filtered with a moving mean filter of width 3. IBI signals were further filtered to remove ectopic beats and strong nonlinearities with a moving median filter of width 3 and Butterworth bandpass filter of order 6 with low-cut and high-cut frequencies of 0.0021 and 0.43 Hz. Signals were divided into 45-minute segments. Features were calculated using a sliding time frame with 50% overlap | A naive bayes classifier achieved an AUROC of 0.82 up to 24 hours before clinical suspicion of sepsis. Adding respiratory signals improved the performance compared with only using heart rate features |
| Simulated and real data (NICU name NR) | Masoudi et al [71], 2013 | Participants; n=32; data size: 233 episodes, ~7 seconds each; model: diagnostic (+2.32-second delay); outcome metric: apnea-bradycardia | 2-channel ECG | No preprocessing techniques were reported. Signals were sampled in 7-second intervals | A coupled HMM model achieved 84.92% sensitivity, 94.17% specificity with a time detection delay of 2.32 (SD 4.82) seconds in detection apnea-bradycardia episodes |
| Simulated and real signals from NICU (NICU name NR) | Altuve et al [72], 2015 | Participants; n=32; data size: 148 RR intervals with a mean duration of 26.25 (SD 11.37) minutes; model: diagnostic (+1.73-second delay); Outcome metric: apnea-bradycardia | ECG (sampling rate NR) | Hidden semi-Markov models to represent the temporal evolution of RR intervals. A preprocessing method that includes quantization and a delayed version of the observation vector is proposed. RR time series was resampled at 10 Hz and segmented at a 7-second interval | The proposed model achieved up to 93.84 (SD 0.79) in specificity and 89.66 (SD 0.71) in sensitivity with a detection delay of 1.59 (SD 0.24) seconds |

| Data set used | Author, year | Study settings | Physiological signal analyzed | Signal processing and computational techniques | Key findings |
|---|---|---|---|---|---|
| NICU (name NR) | Honoré et al [73], 2020 | Participants; n=22; data size: 3501time series, 1200 samples in each; model: prognostic (+72 hours); outcome metric: sepsis | $SpO_2$, respiratory frequency, and RR interval from ECG at 1 Hz | Data were segmented into 20-minute time frames. Time frames with missing data were discarded | Features were extracted from all signals. A combined GMM-HMM model achieved 0.74% (SD 0.05%) accuracy in detecting sepsis. The model was compared with HeRO model, which underperformed using this data set |

[a]CHIME: Collaborative Home Infant Monitoring Evaluation.
[b]NR: not reported.
[c]ECG: electrocardiogram.
[d]$SpO_2$: oxygen saturation.
[e]PICS: Preterm Infant Cardio-Respiratory Signals.
[f]AUROC: area under receiver operating characteristic curve.
[g]FPR: false positive rate.
[h]HRV: heart rate variability.
[i]MIMIC-III: Medical Information Mart for Intensive Care.
[j]HR: heart rate.
[k]SBP: systolic blood pressure.
[l]DBP: diastolic blood pressure.
[m]MBP: mean blood pressure.
[n]LSTM: convolutional neural network-Long Short-Term Memory Network.
[o]LOS: length of stay.
[p]KNN: k-nearest neighbor.
[q]RMSE: root mean square error.
[r]SG: Savitzky-Golay.
[s]CHMM: coupled Hidden Markov Model.
[t]RNN: recurrent neural network.
[u]HMM: Hidden Markov Model.
[v]EEG: electroencephalography.
[w]SVM: support vector machine.
[x]HIE: hypoxic-ischemic encephalopathy.
[y]MAP: mean arterial pressure.
[z]NICU: neonatal intensive care.
[aa]CI: chest impedance.
[ab]DT-CWT: Discrete Time Continuous Wavelet Transform.
[ac]IBI: interbreath variable.
[ad]SII: Signal Instability Index.
[ae]RRE: ribcage respiratory effort.
[af]PR: pulse oximeter.
[ag]HSLDS: Hierarchical Switching Linear Dynamical System.
[ah]RF: random forest.
[ai]GMM: Gaussian Mixture Model.
[aj]FIR: Finite impulse response.
[ak]LMS: least-mean-square.
[al]RIP: respiratory inductive plethysmograph.
[am]ML: machine learning.
[an]IVH: intraventricular hemorrhage.
[ao]BPD: bronchopulmonary dysplasia
[ap]ROP: retinopathy of prematurity.
[aq]NEC: necrotizing enterocolitis

[ar]POPS: pulse oximetry predictive score.
[as]HRC: heart rate characteristics.
[at]HeRO: heart rate observation.
[au]AUC: area under the curve.
[av]ART-M: arterial mean blood pressure.
[aw]NIBP-M: noninvasive blood pressure.
[ax]PPG: photoplethysmography.

## Preprocessing Steps

### *Overview*

Preprocessing of physiological data typically involves several steps, including the handling of missing data, filtering, segmentation, and waveform analysis for feature extraction. Here, we define 5 required preprocessing steps (based on the steps outlined in Berkaya et al [13]) and identify the steps reported by each of the studies in this review (Table 3). The definition of each of the steps is given in subsequent sections.

**Table 3.** Required physiological signal preprocessing steps reported by each of the studies in this review.

| Author, year | Required preprocessing step reported | | | | |
|---|---|---|---|---|---|
| | Handling of missing data | Artifact removal | Resampling, normalization | Waveform feature extraction | Data segmentation |
| Cohen and de Chazal [22], 2013 | ✓ | ✓ | | ✓ | ✓ |
| Cohen and de Chazal [23], 2014 | ✓ | ✓ | ✓ | ✓ | ✓ |
| Cohen and de Chazal [24], 2015 | ✓ | ✓ | | ✓ | ✓ |
| Gee et al [26], 2016 | | | | ✓ | ✓ |
| Gee et al [25], 2017 | | ✓ | | ✓ | ✓ |
| Das et al [27], 2019 | | ✓ | | ✓ | ✓ |
| Mahmud [28], 2019 | | | | | |
| Gee et al [29], 2019 | | ✓ | ✓ | ✓ | ✓ |
| Song et al [30], 2020 | ✓ | ✓ | | N/A[a] | |
| Baker et al [31], 2021 | | ✓ | ✓ | N/A | ✓ |
| Juraev et al [32], 2022 | ✓ | | ✓ | N/A | |
| Montazeri Ghahjaverestan et al [33], 2015 | | ✓ | ✓ | ✓ | |
| Navarro et al [34], 2015 | | ✓ | ✓ | ✓ | ✓ |
| Montazeri Ghahjaverestan et al [35], 2016 | | ✓ | ✓ | ✓ | ✓ |
| León et al [36], 2021 | | | | ✓ | ✓ |
| León et al [37], 2021 | | | | ✓ | ✓ |
| Doyen et al [38], 2021 | | | | ✓ | ✓ |
| Sadoughi et al [39], 2021 | | ✓ | ✓ | ✓ | |
| Ahmed et al [40], 2015 | ✓ | ✓ | ✓ | ✓ | ✓ |
| Temko et al [41], 2015 | | ✓ | ✓ | ✓ | ✓ |
| Lloyd et al [42], 2016 | | ✓ | | N/A | ✓ |
| Semenova et al [43], 2018 | | ✓ | | ✓ | ✓ |
| Semenova et al [44], 2019 | | ✓ | ✓ | ✓ | ✓ |
| Joshi et al [45], 2020 | | ✓ | | ✓ | ✓ |
| Varisco et al [46], 2021 | | ✓ | ✓ | ✓ | |
| Cabrera-Quiros et al [47], 2021 | | | | ✓ | ✓ |
| Varisco et al [48], 2022 | | ✓ | ✓ | ✓ | ✓ |
| Peng et al [49], 2022 | | | | ✓ | ✓ |
| Peng et al [50], 2022 | | ✓ | | ✓ | ✓ |
| Stanculescu et al [51], 2014 | ✓ | ✓ | | N/A | |
| Stanculescu et al [52], 2014 | ✓ | ✓ | | N/A | |
| Shirwaikar et al [53], 2016 | ✓ | ✓ | ✓ | N/A | ✓ |
| Shirwaikar et al [54], 2019 | ✓ | | | N/A | ✓ |
| Williamson et al [55], 2013 | | | ✓ | | ✓ |
| Schiavenato et al [56], 2013 | ✓ | ✓ | | ✓ | |
| Robles-Rubio et al [57], 2014 | | ✓ | ✓ | N/A | |
| Amperayani et al [58], 2017 | | | | | |
| Hu et al [59], 2018 | ✓ | ✓ | ✓ | N/A | ✓ |
| Sullivan et al [60], 2018 | ✓ | | | N/A | ✓ |

| Author, year | Required preprocessing step reported | | | | |
|---|---|---|---|---|---|
| | Handling of missing data | Artifact removal | Resampling, normalization | Waveform feature extraction | Data segmentation |
| Zimmet et al [61], 2020 | ✓ | | | N/A | ✓ |
| Kota et al [62], 2020 | | ✓ | | ✓ | ✓ |
| Mirnia et al [63], 2021 | | | | N/A | |
| Lee et al [64], 2021 | ✓ | ✓ | ✓ | N/A | ✓ |
| Sullivan et al [65], 2021 | ✓ | ✓ | | N/A | ✓ |
| Feng et al [66], 2021 | ✓ | ✓ | ✓ | N/A | ✓ |
| Zuzarte et al [67], 2021 | | ✓ | ✓ | ✓ | ✓ |
| Niestroy et al [68], 2022 | | | | N/A | ✓ |
| Kausch et al [69], 2023 | ✓ | ✓ | | N/A | ✓ |
| Honoré et al [70], 2023 | ✓ | ✓ | ✓ | N/A | ✓ |
| Masoudi et al [71], 2013 | | | | | ✓ |
| Altuve et al [72], 2015 | | ✓ | ✓ | | ✓ |
| Honoré et al [73], 2020 | ✓ | | | N/A | ✓ |

[a]N/A: Not applicable.

### *Handling of Missing Data*

During neonatal physiological monitoring, instances of missing data may arise due to sensor disconnection, improper placements, or signal dropouts. To tackle this issue, methodologies like data imputation or interpolation are applied. For example, if gaps exist in a neonate's HR monitoring data, interpolation methods can estimate the missing values by considering neighboring data points. Widely used interpolation techniques include linear interpolation, spline interpolation, and time-based interpolation. In addition, common data imputation methods involve forward fill, backward fill, and imputation using mean or median values. Methods such as forward fill [30], moving average [44], mean imputation [64,66], and interpolation [67] were used by some studies reviewed in this paper.

### *Artifact Removal*

Neonatal signals can be affected by artifacts, such as those from muscle movements or electrical interference. Commonly used techniques, such as bandpass or notch filters, along with moving averages, are used to effectively eliminate these disturbances. For instance, in neonatal EEG signals, adaptive filters prove beneficial in eliminating artifacts caused by muscle movements, resulting in a clearer representation of the baby's brain activity. Some methods used by the reviewed papers were high-pass filter [27,46] bandpass filter [29,33,44,45,56].

### *Resampling and Normalization*

#### Overview

Resampling is a technique that standardizes data intervals, involving either upsampling (increasing data point frequency) or downsampling (decreasing frequency) to create a regular time series. This aligns signals from different devices or physiological sources. Normalization ensures uniformity and reliability across these standardized sampling rates. For instance, if neonatal HR signals from different devices have varied sampling rates, resampling achieves a common rate, while normalization, using techniques such as minimum-maximum, *z* score, or log scale, ensures consistent amplitude scaling for accurate comparative analysis. In the reviewed studies, normalization techniques such as minimum-maximum [53] and 0 mean normalization [29,59] were used. In terms of resampling, both downsampling [33,34,41] and upsampling [39] techniques were used.

#### Waveform Feature Extraction

Extracting relevant features from a signal's waveform is a fundamental step in signal preprocessing. This involves identifying key characteristics such as peaks, troughs, or other significant points in the signal. In the context of neonatal ECG, feature extraction may involve identifying key points such as R-peaks to analyze HRV, providing valuable insights into the infant's autonomic nervous system development. The Pan-Tompkins algorithm is a popular method chosen by multiple papers reviewed in this study that conducted R-peak detection from the QRS complex [22,24,27,33,35,39].

#### Data Segmentation

Segmenting data is the process of breaking down a continuous signal into smaller, more manageable sections to enable targeted analysis. This practice is especially beneficial when dealing with lengthy signals. Data segmentation is a common preprocessing step in ML workflows. For instance, in the analysis of neonatal sleep patterns using EEG, data segmentation can involve dividing the continuous EEG signal into epochs, allowing for the identification and study of sleep stages in shorter, more manageable segments. Commonly used segmentation techniques include fixed length, sliding window, and threshold- and feature-based segmentation. Some of the data segmentation sizes used in the reviewed studies were 30-second [22-24,45] and 1-minute [41] epochs and a sliding window of varied sizes [35,40,55,59,64].

In neonatal physiological signal processing, these preprocessing techniques contribute to the accurate interpretation of signals, aiding health care professionals in monitoring and providing appropriate care in the NICU or other clinical settings.

It can be seen from Table 3 that only 7 (13%) out of the 52 reviewed studies reported all the recommended preprocessing steps. This could have several impacts on the downstream analysis. For instance, several papers missed reporting on how they segmented the data for feature extraction and classification, although it is essential for clinical validation in cases where the segment duration is dependent on the adverse outcome prediction performance. In HRV analysis, it is important to indicate whether it is a short-term (~5 minutes) or a long-term (≥24 hours) analysis as they reflect different underlying physiological processes and thus demonstrate different predictive power [107]. Along with the segment duration, additional information such as the sampling rate of the signals will provide a clear reflection of the data set size. Downsampling the data to a low sampling rate (eg, 50 Hz) has also shown a significant impact on HRV analysis [108]. Although all the reviewed studies mentioned the participant number, and majority of them (n=39) reported the sampling rate of the signals, very few provided details on the sample size or data set duration or whether the data set was resampled for subsequent analysis. These elements provide a clearer picture of the computational time and resources required for clinical validation and adoption. Although physiological recordings collected in the NICU environment suffer greatly from missing data due to similar factors that introduce artifacts [109], reporting how missing data are handled is scarce. Different methods for dealing with missing values could cause different results, and not all might be suitable for a particular problem. Therefore, it is important to report all the details related to the adopted approach.

The incomplete or partial reporting found in these studies has significant implications for the implementation of QMS in using these techniques for clinical adoption. A good implementation of QMS requires a comprehensive reporting of each intermediary step involved in constructing an AI and ML pipeline. The International Medical Device Regulators Forum offers guidance on the clinical evaluation required for any product intended for use as a medical device [110]. According to the International Medical Device Regulators Forum guidelines, during clinical evaluation, relevant research articles are reviewed to identify clinical evidence supporting the product [111]. The guideline encourages manufacturers to follow these recognized standards and best practices in the development, validation, and manufacturing processes. Clinical evaluations are required by the European Union medical device regulation, and it is also mentioned in the ISO 13485 (the quality management standard for medical devices). Thus, detailed reporting is crucial as it can be used by regulatory bodies to evaluate future SaMD products clinically. Steps such as the missing data handling procedures are also required by the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) checklist for model development and validation, which assesses the risk of bias and clinical usefulness of the prediction model [112]. Another example is a questionnaire prepared by the German Notified Body Interest Group, and it was adopted to assess some AI-powered medical products in the European Union. This questionnaire includes inquiries about data management, including data collection, labeling, preprocessing procedures, and relevant documentation. Transparent and detailed reporting of these steps is essential to ensure the safety, efficacy, and reliability of SaMD.

## Discussion

### Principal Findings

This review aimed to summarize the computational methods used for preprocessing preterm infants' physiological data as a first step in developing data-driven predictive models for adverse outcomes related to clinical decision support. This is an important step, especially from a clinician's perspective, because it increases the trustworthiness of the developed models by allowing for the verification and reproduction of the results. In addition, it aids in achieving regulatory compliance and ensures the safety, efficacy, and ethical use of AI-based health care devices. Furthermore, it allows us to recognize the shortcomings in the current state-of-the-art studies and recommend guidelines for transparent reporting. The review found that the studies were heterogeneous in terms of their methods and applications. Therefore, a narrative approach to reporting the results was taken instead of a quantitative approach. Through the analysis we identified several key components that were incomplete or partially reported by the included studies, which are summarized in Table 3. To ensure transparent reporting for any future studies in this area, we recommend detailed reporting of all preprocessing steps listed in Table 3, which will allow revealing their strengths and weaknesses and ultimately make them usable and reproducible. Reproducible research allows clinicians to make more informed decisions about patient care and treatment based on the evidence that has been thoroughly assessed.

### Comparison With Prior Work

The reviews published in recent years have highlighted the potential of big data and AI in supporting clinical decision-making in the neonatal health care domain [10,15,21,113,114], particularly in using physiological data for detecting or predicting neonatal health outcomes. However, appropriate preprocessing of these data is a prerequisite for developing clinically deployable models. A systematic review by McAdams et al [10] reported different ML models used to predict different clinical outcomes in neonates. However, their primary focus was on 5 neonatal morbidities, and they did not focus on reporting the preprocessing methods applied before building the ML models. Furthermore, they did not include studies using real-time continuous physiological data; 28 out of their 68 studies were based on physiological data (not continuous), and the rest were based on electronic medical records and imaging data. Latremouille et al [15] performed a review on HRV analysis for neonates. The primary limitation of the work was the lack of reporting in detail about the preprocessing steps of ECG signals before HRV analysis, such as ECG handling and segmentation, R-wave (QRS complex) identification technique, software and parameters, and ranges of all HRV features. They identified these components as

incomplete or missing in the studies they reviewed and thus recommended clear reporting of these aspects for future studies in this area. These limitations served as a motivation for our review to focus on the preprocessing techniques of neonatal physiological signals in a broader sense, which serves as the preliminary step for any big data–based approaches.

### Limitations

There are several limitations to this review. Screening of all the included studies was conducted independently by 1 reviewer, which may have introduced bias. In addition, this review did not include a quantitative or comparative analysis of the reviewed studies, as the techniques used to analyze the physiological signals were diverse. Future work could include a quantitative evaluation of the studies that were homogeneous in design.

### Conclusions

This review explores the computational methods used by the current state-of-the-art ML-driven clinical decision support approaches to preprocess physiological signals collected from infants treated in the neonatal setting. A summary of the studies identified heterogeneity in the techniques used for analysis and revealed a lack of consistent and detailed reporting, which is important for building robust, transparent, and clinically deployable prediction models. The availability of powerful hardware and software resources in the NICU environment and growing interest in big data and AI are driving strong demand for clinical decision support applications. We recommend clear reporting of the different steps in the preprocessing of the neonatal physiological signals to ensure transparency in clinical validation and accelerate the adoption of developed models in the clinical setting. This will further enhance the delivery and adoption of reliable, regulatory-compliant, safe, and effective products in health care.

## Acknowledgments

The work is supported by the Cerebral Palsy Alliance and the University of Sydney (ethics approval number 2021/ETH00824).

## Data Availability

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

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search queries.
[DOCX File , 18 KB - ijmr_v13i1e46946_app1.docx ]

Multimedia Appendix 2
Bibliography files for all databases.
[ZIP File (Zip Archive), 2585 KB - ijmr_v13i1e46946_app2.zip ]

Multimedia Appendix 3
All included papers.
[ZIP File (Zip Archive), 61626 KB - ijmr_v13i1e46946_app3.zip ]

Multimedia Appendix 4
Metadata of all papers.
[XLSX File (Microsoft Excel File), 2654 KB - ijmr_v13i1e46946_app4.xlsx ]

Multimedia Appendix 5
Metadata of all papers in the full-text review.
[XLSX File (Microsoft Excel File), 78 KB - ijmr_v13i1e46946_app5.xlsx ]

Multimedia Appendix 6
PRISMA-ScR checklist.
[DOCX File , 84 KB - ijmr_v13i1e46946_app6.docx ]

## References

1. Platt MJ. Outcomes in preterm infants. Public Health 2014 May;128(5):399-403 [FREE Full text] [doi: 10.1016/j.puhe.2014.03.010] [Medline: 24794180]
2. Seaton SE, Barker L, Draper ES, Abrams KR, Modi N, Manktelow BN. Estimating neonatal length of stay for babies born very preterm. Arch Dis Child Fetal Neonatal Ed 2019 Mar;104(2):F182-F186 [FREE Full text] [doi: 10.1136/archdischild-2017-314405] [Medline: 29588296]
3. De Georgia MA, Kaffashi F, Jacono FJ, Loparo KA. Information technology in critical care: review of monitoring and data acquisition systems for patient care and research. ScientificWorldJournal 2015;2015:727694 [FREE Full text] [doi: 10.1155/2015/727694] [Medline: 25734185]
4. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. IEEE J Biomed Health Inform 2015 Jul;19(4):1193-1208 [FREE Full text] [doi: 10.1109/JBHI.2015.2450362] [Medline: 26173222]
5. Kumar N, Akangire G, Sullivan B, Fairchild K, Sampath V. Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront. Pediatr Res 2020 Jan;87(2):210-220 [FREE Full text] [doi: 10.1038/s41390-019-0527-0] [Medline: 31377752]
6. Baljak V, Ljubovic A, Michel J, Montgomery M, Salaway R. A scalable realtime analytics pipeline and storage architecture for physiological monitoring big data. Smart Health 2018 Dec;9-10:275-286 [FREE Full text] [doi: 10.1016/j.smhl.2018.07.013]
7. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 03;157(1):29-43 [FREE Full text] [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
8. Comito C, Falcone D, Forestiero A. Current trends and practices in smart health monitoring and clinical decision support. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. 2020 Presented at: BIBM 2020; December 16-19, 2020; Seoul, Korea (South). [doi: 10.1109/BIBM49941.2020.9313449]
9. Yu KH, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019 Mar;28(3):238-241. [doi: 10.1136/bmjqs-2018-008551] [Medline: 30291179]
10. McAdams RM, Kaur R, Sun Y, Bindra H, Cho SJ, Singh H. Predicting clinical outcomes using artificial intelligence and machine learning in neonatal intensive care units: a systematic review. J Perinatol 2022 Dec;42(12):1561-1575. [doi: 10.1038/s41372-022-01392-8] [Medline: 35562414]
11. Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance. Med Phys 2020 Jun;47(5):e228-e235 [FREE Full text] [doi: 10.1002/mp.13562] [Medline: 32418341]
12. Sloane EJ, Silva RJ. Chapter 83 - Artificial intelligence in medical devices and clinical decision support systems. In: Iadanza E, editor. Clinical Engineering Handbook. Cambridge, MA: Academic Press; 2019:556-568.
13. Kaplan Berkaya S, Uysal AK, Sora Gunal E, Ergin S, Gunal S, Gulmezoglu MB. A survey on ECG analysis. Biomed Signal Process Control 2018 May;43:216-235 [FREE Full text] [doi: 10.1016/j.bspc.2018.03.003]
14. Sweeney KT, Ward TE, McLoone SF. Artifact removal in physiological signals—practices and possibilities. IEEE Trans Inform Technol Biomed 2012 May;16(3):488-500 [FREE Full text] [doi: 10.1109/titb.2012.2188536]
15. Latremouille S, Lam J, Shalish W, Sant'Anna G. Neonatal heart rate variability: a contemporary scoping review of analysis methods and clinical applications. BMJ Open 2021 Dec 21;11(12):e055209 [FREE Full text] [doi: 10.1136/bmjopen-2021-055209] [Medline: 34933863]
16. Baird P, Cobbaert K. Software as a medical device: a comparison of the EU's approach with the US's approach. British Standards Institution. URL: https://www.bsigroup.com/globalassets/meddev/localfiles/fr-fr/whitepapers/md-wp-software.pdf [accessed 2024-07-24]
17. Simera I. Get the content right: following reporting guidelines will make your research paper more complete, transparent and usable. J Pak Med Assoc 2013 Feb;63(2):283-285. [Medline: 23894916]
18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
19. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. York, UK: Centre for Reviews and Dissemination; 2009.
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
21. Mangold C, Zoretic S, Thallapureddy K, Moreira A, Chorath K, Moreira A. Machine learning models for predicting neonatal mortality: a systematic review. Neonatology 2021;118(4):394-405 [FREE Full text] [doi: 10.1159/000516891] [Medline: 34261070]
22. Cohen G, de Chazal P. Detection of sleep apnoea in infants using ECG and oximetry signals. In: Proceedings of the Computing in Cardiology 2013. 2013 Presented at: CinC 2013; September 22-25, 2013; Zaragoza, Spain.
23. Cohen G, de Chazal P. A multi-modal approach to sleep-wake classification in infants using minimally invasive sensors. In: Proceedings of the Computing in Cardiology. 2014 Presented at: CinC 2014; September 7-10, 2014; Cambridge, MA.

24. Cohen G, de Chazal P. Automated detection of sleep apnea in infants: a multi-modal approach. Comput Biol Med 2015 Aug;63:118-123 [FREE Full text] [doi: 10.1016/j.compbiomed.2015.05.007] [Medline: 26073098]
25. Gee AH, Barbieri R, Paydarfar D, Indic P. Predicting bradycardia in preterm infants using point process analysis of heart rate. IEEE Trans Biomed Eng 2017 Sep;64(9):2300-2308 [FREE Full text] [doi: 10.1109/tbme.2016.2632746]
26. Gee AH, Barbieri R, Paydarfar D, Indic P. Improving heart rate estimation in preterm infants with bivariate point process analysis of heart rate and respiration. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 Presented at: EMBC 2016; August 16-20, 2016; Orlando, FL. [doi: 10.1109/embc.2016.7590851]
27. Das S, Moraffah B, Banerjee A, Gupta SK, Papandreou-Suppappola A. Bradycardia prediction in preterm infants using nonparametric kernel density estimation. In: Proceedings of the 53rd Asilomar Conference on Signals, Systems, and Computers. 2019 Presented at: ACSSC 2019; November 3-6, 2019; Pacific Grove, CA. [doi: 10.1109/ieeeconf44664.2019.9049007]
28. Mahmud MS, Wang H, Kim Y. Accelerated prediction of bradycardia in preterm infants using time-frequency analysis. In: Proceedings of the International Conference on Computing, Networking and Communications. 2019 Presented at: ICNC 2019; February 18-21, 2019; Honolulu, HI. [doi: 10.2172/10116953]
29. Gee AH, Garcia-Olano D, Ghosh J, Paydarfar D. Explaining deep classification of time-series data with learned prototypes. CEUR Workshop Proc 2019 Aug;2429:15-22 [FREE Full text] [Medline: 33867901]
30. Song W, Jung SY, Baek H, Choi CW, Jung YH, Yoo S. A predictive model based on machine learning for the early detection of late-onset neonatal sepsis: development and observational study. JMIR Med Inform 2020 Jul 31;8(7):e15965 [FREE Full text] [doi: 10.2196/15965] [Medline: 32735230]
31. Baker S, Xiang W, Atkinson I. Hybridized neural networks for non-invasive and continuous mortality risk assessment in neonates. Comput Biol Med 2021 Jul;134:104521 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104521] [Medline: 34111664]
32. Juraev F, El-Sappagh S, Abdukhamidov E, Ali F, Abuhmed T. Multilayer dynamic ensemble model for intensive care unit mortality prediction of neonate patients. J Biomed Inform 2022 Nov;135:104216 [FREE Full text] [doi: 10.1016/j.jbi.2022.104216] [Medline: 36208833]
33. Montazeri Ghahjaverestan N, Shamsollahi MB, Ge D, Hernández AI. Switching Kalman filter based methods for apnea bradycardia detection from ECG signals. Physiol Meas 2015 Sep;36(9):1763-1783 [FREE Full text] [doi: 10.1088/0967-3334/36/9/1763] [Medline: 26235524]
34. Navarro X, Porée F, Beuchée A, Carrault G. Artifact rejection and cycle detection in immature breathing: application to the early detection of neonatal sepsis. Biomed Signal Processing Control 2015 Feb;16:9-16 [FREE Full text] [doi: 10.1016/j.bspc.2014.10.007]
35. Montazeri Ghahjaverestan N, Masoudi S, Shamsollahi MB, Beuchee A, Pladys P, Ge D, et al. Coupled hidden Markov model-based method for apnea bradycardia detection. IEEE J Biomed Health Inform 2016 Mar;20(2):527-538 [FREE Full text] [doi: 10.1109/JBHI.2015.2405075] [Medline: 25706937]
36. Leon C, Carrault G, Pladys P, Beuchee A. Early detection of late onset sepsis in premature infants using visibility graph analysis of heart rate variability. IEEE J Biomed Health Inform 2021 Apr;25(4):1006-1017 [FREE Full text] [doi: 10.1109/JBHI.2020.3021662] [Medline: 32881699]
37. León C, Pladys P, Beuchée A, Carrault G. Recurrent neural networks for early detection of late onset sepsis in premature infants using heart rate variability. In: Proceedings of the Computing in Cardiology. 2021 Presented at: CinC 2021; September 13-15, 2021; Brno, Czech Republic. [doi: 10.23919/cinc53138.2021.9662715]
38. Doyen M, Hernández AI, Flamant C, Defontaine A, Favrais G, Altuve M, et al. Early bradycardia detection and therapeutic interventions in preterm infant monitoring. Sci Rep 2021 May 18;11(1):10486 [FREE Full text] [doi: 10.1038/s41598-021-89468-x] [Medline: 34006917]
39. Sadoughi A, Shamsollahi MB, Fatemizadeh E, Beuchée A, Hernández AI, Montazeri Ghahjaverestan N. Detection of apnea bradycardia from ECG signals of preterm infants using layered hidden Markov model. Ann Biomed Eng 2021 Sep;49(9):2159-2169 [FREE Full text] [doi: 10.1007/s10439-021-02732-z] [Medline: 33638031]
40. Ahmed R, Temko A, Marnane WP, Boylan G, Lightbody G. Classification of hypoxic-ischemic encephalopathy using long term heart rate variability based features. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2015 Presented at: EMBC 2015; August 25-29, 2015; Milan, Italy. [doi: 10.1109/embc.2015.7318866]
41. Temko A, Doyle O, Murray D, Lightbody G, Boylan G, Marnane W. Multimodal predictor of neurodevelopmental outcome in newborns with hypoxic-ischaemic encephalopathy. Comput Biol Med 2015 Aug;63:169-177 [FREE Full text] [doi: 10.1016/j.compbiomed.2015.05.017] [Medline: 26093065]
42. Lloyd RO, O'Toole JM, Livingstone V, Hutch WD, Pavlidis E, Cronin AM, et al. Predicting 2-y outcome in preterm infants using early multimodal physiological monitoring. Pediatr Res 2016 Sep;80(3):382-388 [FREE Full text] [doi: 10.1038/pr.2016.92] [Medline: 27089498]
43. Semenova O, Carra G, Lightbody G, Boylan G, Dempsey E, Temko A. Heart rate variability during periods of low blood pressure as a predictor of short-term outcome in preterms. In: Proceedings of the 40th Annual International Conference of

the IEEE Engineering in Medicine and Biology Society. 2018 Presented at: EMBC 2018; July 18-21, 2018; Honolulu, HI. [doi: 10.1109/embc.2018.8513600]
44. Semenova O, Carra G, Lightbody G, Boylan G, Dempsey E, Temko A. Prediction of short-term health outcomes in preterm neonates from heart-rate variability and blood pressure using boosted decision trees. Comput Methods Programs Biomed 2019 Oct;180:104996 [FREE Full text] [doi: 10.1016/j.cmpb.2019.104996] [Medline: 31421605]
45. Joshi R, Kommers D, Oosterwijk L, Feijs L, van Pul C, Andriessen P. Predicting neonatal sepsis using features of heart rate variability, respiratory characteristics, and ECG-derived estimates of infant motion. IEEE J Biomed Health Inform 2020 Mar;24(3):681-692 [FREE Full text] [doi: 10.1109/JBHI.2019.2927463] [Medline: 31295130]
46. Varisco G, Kommers D, Long X, Zhan Z, Nano MM, Cottaar W, et al. Optimized detection of central apneas preceding late-onset sepsis in premature infants. In: Proceedings of the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2021 Presented at: EMBC 2021; November 1-5, 2021; Virtual Event. [doi: 10.1109/embc46164.2021.9629528]
47. Cabrera-Quiros L, Kommers D, Wolvers MK, Oosterwijk L, Arents N, van der Sluijs-Bens J, et al. Prediction of late-onset sepsis in preterm infants using monitoring signals and machine learning. Crit Care Explor 2021 Jan 27;3(1):e0302 [FREE Full text] [doi: 10.1097/CCE.0000000000000302] [Medline: 33532727]
48. Varisco G, Peng Z, Kommers D, Zhan Z, Cottaar W, Andriessen P, et al. Central apnea detection in premature infants using machine learning. Comput Methods Programs Biomed 2022 Nov;226:107155 [FREE Full text] [doi: 10.1016/j.cmpb.2022.107155] [Medline: 36215858]
49. Peng Z, Varisco G, Liang RH, Kommers D, Cottaar W, Andriessen P, et al. DeepLOS: deep learning for late-onset sepsis prediction in preterm infants using heart rate variability. Smart Health 2022 Dec;26:100335 [FREE Full text] [doi: 10.1016/j.smhl.2022.100335]
50. Peng Z, Varisco G, Long X, Liang RH, Kommers D, Cottaar W, et al. A continuous late-onset sepsis prediction algorithm for preterm infants using multi-channel physiological signals from a patient monitor. IEEE J Biomed Health Inform 2023 Jan;27(1):550-561. [doi: 10.1109/JBHI.2022.3216055] [Medline: 36264730]
51. Stanculescu I, Williams CK, Freer Y. Autoregressive hidden Markov models for the early detection of neonatal sepsis. IEEE J Biomed Health Inform 2014 Sep;18(5):1560-1570 [FREE Full text] [doi: 10.1109/JBHI.2013.2294692] [Medline: 25192568]
52. Stanculescu I, Williams CK, Freer Y. A hierarchical switching linear dynamical system applied to the detection of sepsis in neonatal condition monitoring. In: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence. 2014 Presented at: UAI'14; July 23-27, 2014; Quebec City, Quebec.
53. Shirwaikar RD, Acharya UD, Makkithaya K, Surulivelr M, Lewis LE. Machine learning techniques for neonatal apnea prediction. J Artif Intell 2016;9(1-3):33-38 [FREE Full text] [doi: 10.3923/jai.2016.33.38]
54. Shirwaikar RD, Acharya UD, Makkithaya K, Srivastava S, Lewis UL, M S. Optimizing neural networks for medical data sets: a case study on neonatal apnea prediction. Artif Intell Med 2019 Jul;98:59-76 [FREE Full text] [doi: 10.1016/j.artmed.2019.07.008] [Medline: 31521253]
55. Williamson JR, Bliss DW, Browne DW, Indic P, Bloch-Salisbury E, Paydarfar D. Individualized apnea prediction in preterm infants using cardio-respiratory and movement signals. In: Proceedings of the IEEE International Conference on Body Sensor Networks. 2013 Presented at: BSN 2013; May 6-9, 2013; Cambridge, MA. [doi: 10.1109/BSN.2013.6575523]
56. Schiavenato M, Oliu C, Bello E, Bohorquez J, Claure N. Development of a system for the assessment of heart rate variability in the NICU. In: Proceedings of the 29th Southern Biomedical Engineering Conference. 2013 Presented at: SBEC 2013; May 3-5, 2013; Miami, FL. [doi: 10.1109/SBEC.2013.21]
57. Robles-Rubio CA, Brown KA, Bertolizio G, Kearney RE. Automated analysis of respiratory behavior for the prediction of apnea in infants following general anesthesia. In: Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014 Presented at: EMBC 2014; August 26-30, 2014; Chicago, IL. [doi: 10.1109/embc.2014.6943579]
58. Amperayani VA, Indic P, Travers CP, Barbieri R, Paydarfar D, Ambalavanan N. An algorithm for risk stratification of preterm infants. In: Proceedings of the Computing in Cardiology. 2017 Presented at: CinC 2017; September 24-27, 2017; Rennes, France. [doi: 10.22489/CinC.2017.333-127]
59. Hu Y, Lee VC, Tan K. Prediction of clinicians' treatment in preterm infants with suspected late-onset sepsis — an ML approach. In: Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications. 2018 Presented at: ICIEA 2018; May 31-June 2, 2018; Wuhan, China. [doi: 10.1109/ICIEA.2018.8397888]
60. Sullivan BA, Wallman-Stokes A, Isler J, Sahni R, Moorman JR, Fairchild KD, et al. Early pulse oximetry data improves prediction of death and adverse outcomes in a two-center cohort of very low birth weight infants. Am J Perinatol 2018 Nov;35(13):1331-1338 [FREE Full text] [doi: 10.1055/s-0038-1654712] [Medline: 29807371]
61. Zimmet AM, Sullivan BA, Moorman JR, Lake DE, Ratcliffe SJ. Trajectories of the heart rate characteristics index, a physiomarker of sepsis in premature infants, predict Neonatal ICU mortality. JRSM Cardiovasc Dis 2020 Aug 05;9:2048004020945142 [FREE Full text] [doi: 10.1177/2048004020945142] [Medline: 33240492]

62. Kota S, du Plessis A, Massaro AN, Chang T, Al-Shargabi T, Govindan RB. A frequency based spatial filter to mitigate volume conduction in electroencephalogram signals. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:4001-4004. [doi: 10.1109/EMBC.2016.7591604] [Medline: 28269162]
63. Mirnia K, Heidarzadeh M, Afjeh SA, Alizadeh P, Kashan AA, Bordbar A, et al. Signal processing of heart rate for predicting sepsis in premature neonates. J Med Signals Sens 2021 Jul 21;11(3):222-226 [FREE Full text] [doi: 10.4103/jmss.JMSS_30_20] [Medline: 34466402]
64. Lee J, Cai J, Li F, Vesoulis ZA. Predicting mortality risk for preterm infants using random forest. Sci Rep 2021 Mar 31;11(1):7308 [FREE Full text] [doi: 10.1038/s41598-021-86748-4] [Medline: 33790395]
65. Sullivan BA, Nagraj VP, Berry KL, Fleiss N, Rambhia A, Kumar R, et al. Clinical and vital sign changes associated with late-onset sepsis in very low birth weight infants at 3 NICUs. J Neonatal Perinatal Med 2021;14(4):553-561 [FREE Full text] [doi: 10.3233/NPM-200578] [Medline: 33523025]
66. Feng J, Lee J, Vesoulis ZA, Li F. Predicting mortality risk for preterm infants using deep learning models with time-series vital sign data. NPJ Digit Med 2021 Jul 14;4(1):108 [FREE Full text] [doi: 10.1038/s41746-021-00479-4] [Medline: 34262112]
67. Zuzarte I, Sternad D, Paydarfar D. Predicting apneic events in preterm infants using cardio-respiratory and movement features. Comput Methods Programs Biomed 2021 Sep;209:106321 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106321] [Medline: 34380078]
68. Niestroy JC, Moorman JR, Levinson MA, Manir SA, Clark TW, Fairchild KD, et al. Discovery of signatures of fatal neonatal illness in vital signs using highly comparative time-series analysis. NPJ Digit Med 2022 Jan 17;5(1):6 [FREE Full text] [doi: 10.1038/s41746-021-00551-z] [Medline: 35039624]
69. Kausch SL, Brandberg JG, Qiu J, Panda A, Binai A, Isler J, et al. Cardiorespiratory signature of neonatal sepsis: development and validation of prediction models in 3 NICUs. Pediatr Res 2023 Jun;93(7):1913-1921 [FREE Full text] [doi: 10.1038/s41390-022-02444-7] [Medline: 36593281]
70. Honoré A, Forsberg D, Adolphson K, Chatterjee S, Jost K, Herlenius E. Vital sign-based detection of sepsis in neonates using machine learning. Acta Paediatr 2023 Apr;112(4):686-696 [FREE Full text] [doi: 10.1111/apa.16660] [Medline: 36607251]
71. Masoudi S, Montazeri N, Shamsollahi MB, Ge D, Beuchée A, Pladys P, et al. Early detection of apnea-bradycardia episodes in preterm infants based on coupled hidden Markov model. In: Proceedings of the IEEE International Symposium on Signal Processing and Information Technology. 2013 Presented at: ISSPIT 2013; December 12-15, 2013; Athens, Greece. [doi: 10.1109/ISSPIT.2013.6781887]
72. Altuve M, Carrault G, Beuchée A, Pladys P, Hernández AI. Online apnea-bradycardia detection based on hidden semi-Markov models. Med Biol Eng Comput 2015 Jan;53(1):1-13 [FREE Full text] [doi: 10.1007/s11517-014-1207-1] [Medline: 25300402]
73. Honoré A, Liu D, Forsberg D, Coste K, Herlenius E, Chatterjee S, et al. Hidden Markov models for sepsis detection in preterm infants. In: Proceedings of the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. 2020 Presented at: ICASSP 2020; May 4-8, 2020; Barcelona, Spain. [doi: 10.1109/icassp40776.2020.9054635]
74. Continuous patient monitors. Phillips Healthcare. URL: https://www.philips.com.au/healthcare/solutions/patient-monitoring/continuous-patient-monitoring-systems [accessed 2024-07-21]
75. BedMasterEx. Anandic. URL: https://www.anandic.com/produkte/produkt-detail/bedmasterex/ [accessed 2024-07-21]
76. NicoletOne™ EEG system. Natus. URL: https://natus.com/neuro/nicoletone-eeg-system/ [accessed 2024-07-21]
77. Patient Information Center iX (PIC iX). Philips. URL: https://www.philips.com.au/healthcare/product/HCNOCTN171/patient-information-center-ix-pic-ix [accessed 2024-07-21]
78. Mann KD, Good NM, Fatehi F, Khanna S, Campbell V, Conway R, et al. Predicting patient deterioration: a review of tools in the digital hospital setting. J Med Internet Res 2021 Sep 30;23(9):e28209 [FREE Full text] [doi: 10.2196/28209] [Medline: 34591017]
79. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Eur Heart J 1996 Mar 01;17(3):354-381 [FREE Full text] [doi: 10.1093/oxfordjournals.eurheartj.a014868]
80. Paliwoda M, New K, Davies M, Bogossian F. Physiological vital sign ranges in newborns from 34 weeks gestation: a systematic review. Int J Nurs Stud 2018 Jan;77:81-90 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.10.004] [Medline: 29055243]
81. Gardner SL, Carter BS, Enzman-Hines M, Niermeyer S. Merenstein & Gardner's Handbook of Neonatal Intensive Care: An Interprofessional Approach. Amsterdam, The Netherlands: Elsevier; 2020.
82. Sinha S, Miall L, Jardine L. Essential Neonatal Medicine. Hoboken, NJ: John Wiley & Sons; 2012.
83. Paliwoda M, New K, Bogossian F, Ballard E. Physiological vital sign reference ranges for well late preterm newborns calculated during a typical 2 h newborn period between 2 h and 7 days of life. Physiol Meas 2021 Sep 27;42(9):095001 [FREE Full text] [doi: 10.1088/1361-6579/ac155b] [Medline: 34271562]

84. Crowell DH, Brooks LJ, Colton T, Corwin MJ, Hoppenbrouwers TT, Hunt CE, et al. Infant polysomnography: reliability. Collaborative Home Infant Monitoring Evaluation (CHIME) Steering Committee. Sleep 1997 Jul;20(7):553-560. [Medline: 9322271]
85. Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985 Mar;BME-32(3):230-236 [FREE Full text] [doi: 10.1109/tbme.1985.325532]
86. de Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P, O'Malley M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. IEEE Trans Biomed Eng 2003 Jun;50(6):686-696 [FREE Full text] [doi: 10.1109/tbme.2003.812203]
87. Lewicke AT, Sazonov ES, Schuckers SA. Sleep-wake identification in infants: heart rate variability compared to actigraphy. In: Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2004 Presented at: IEMBS 2004; September 1-5, 2004; San Francisco, CA. [doi: 10.1109/iembs.2004.1403189]
88. Goldberger AL, Amaral L, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet : components of a new research resource for complex physiologic signals. Circulation 2000 Jun 13;101(23) [FREE Full text] [doi: 10.1161/01.CIR.101.23.e215]
89. Torrence C, Compo GP. A practical guide to wavelet analysis. Bull Am Meteorol Soc 1998;79(1):61-78 [FREE Full text] [doi: 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2]
90. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
91. Savitzky A, Golay MJ. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 1964;36(8):1627-1639 [FREE Full text] [doi: 10.1021/ac60214a047]
92. Dumont J, Hernandez AI, Fleureau J, Carrault G. Modelling temporal evolution of cardiac electrophysiological features using Hidden Semi-Markov Models. In: Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2008 Presented at: IEMBS 2008; August 20-25, 2008; Vancouver, BC. [doi: 10.1109/iembs.2008.4649116]
93. Doyen M, Ge D, Beuchée A, Carrault G, Hernández AI. Robust, real-time generic detector based on a multi-feature probabilistic method. PLoS One 2019 Oct 29;14(10):e0223785 [FREE Full text] [doi: 10.1371/journal.pone.0223785] [Medline: 31661497]
94. Hamilton PS, Tompkins WJ. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. IEEE Trans Biomed Eng 1986 Dec;33(12):1157-1165 [FREE Full text] [doi: 10.1109/tbme.1986.325695] [Medline: 3817849]
95. Rooijakkers MJ, Rabotti C, Oei SG, Mischi M. Low-complexity R-peak detection for ambulatory fetal monitoring. Physiol Meas 2012 Jul;33(7):1135-1150 [FREE Full text] [doi: 10.1088/0967-3334/33/7/1135] [Medline: 22735075]
96. Lee H, Rusin CG, Lake DE, Clark MT, Guin L, Smoot TJ, et al. A new algorithm for detecting central apnea in neonates. Physiol Meas 2012 Jan;33(1):1-17 [FREE Full text] [doi: 10.1088/0967-3334/33/1/1] [Medline: 22156193]
97. Mohr MA, Vergales BD, Lee H, Clark MT, Lake DE, Mennen AC, et al. Very long apnea events in preterm infants. J Appl Physiol (1985) 2015 Mar 01;118(5):558-568 [FREE Full text] [doi: 10.1152/japplphysiol.00144.2014] [Medline: 25549762]
98. Vergales BD, Paget-Brown AO, Lee H, Guin LE, Smoot TJ, Rusin CG, et al. Accurate automated apnea analysis in preterm infants. Am J Perinatol 2014 Feb;31(2):157-162 [FREE Full text] [doi: 10.1055/s-0033-1343769] [Medline: 23592319]
99. Redmond SJ, Heneghan C. Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. IEEE Trans Biomed Eng 2006 Mar;53(3):485-496 [FREE Full text] [doi: 10.1109/TBME.2005.869773] [Medline: 16532775]
100. Zuzarte I, Indic P, Sternad D, Paydarfar D. Quantifying movement in preterm infants using photoplethysmography. Ann Biomed Eng 2019 Feb;47(2):646-658 [FREE Full text] [doi: 10.1007/s10439-018-02135-7] [Medline: 30255214]
101. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 1989 Feb;77(2):257-286 [FREE Full text] [doi: 10.1109/5.18626]
102. Lomb NR. Least-squares frequency analysis of unequally spaced data. Astrophys Space Sci 1976 Feb;39(2):447-462 [FREE Full text] [doi: 10.1007/BF00648343]
103. Kota S, Massaro AN, Chang T, Al-Shargabi T, Cristante C, Vezina G, et al. Prognostic value of continuous electroencephalogram delta power in neonates with hypoxic-ischemic encephalopathy. J Child Neurol 2020 Jul;35(8):517-525 [FREE Full text] [doi: 10.1177/0883073820915323] [Medline: 32306827]
104. Govindan RB, Kota S, Al-Shargabi T, Massaro AN, Chang T, du Plessis A. Effect of electrocardiogram interference on cortico-cortical connectivity analysis and a possible solution. J Neurosci Methods 2016 Sep 01;270:76-84 [FREE Full text] [doi: 10.1016/j.jneumeth.2016.06.009] [Medline: 27291356]
105. Govindan RB, Massaro A, Vezina G, Tsuchida T, Cristante C, du Plessis A. Does relative or absolute EEG power have prognostic value in HIE setting? Clin Neurophysiol 2017 Jan;128(1):14-15 [FREE Full text] [doi: 10.1016/j.clinph.2016.10.094] [Medline: 27866114]
106. Halliday DM, Rosenberg JR, Amjad AM, Breeze P, Conway BA, Farmer SF. A framework for the analysis of mixed time series/point process data--theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. Prog Biophys Mol Biol 1995;64(2-3):237-278 [FREE Full text] [doi: 10.1016/s0079-6107(96)00009-0] [Medline: 8987386]

107. Shaffer F, Meehan ZM, Zerr CL. A critical review of ultra-short-term heart rate variability norms research. Front Neurosci 2020 Nov 19;14:594880 [FREE Full text] [doi: 10.3389/fnins.2020.594880] [Medline: 33328866]
108. Kwon O, Jeong J, Kim HB, Kwon IH, Park SY, Kim JE, et al. Electrocardiogram sampling frequency range acceptable for heart rate variability analysis. Healthc Inform Res 2018 Jul;24(3):198-206 [FREE Full text] [doi: 10.4258/hir.2018.24.3.198] [Medline: 30109153]
109. Gravel J, Opatrny L, Gouin S. High rate of missing vital signs data at triage in a paediatric emergency department. Paediatr Child Health 2006 Apr;11(4):211-215 [FREE Full text] [doi: 10.1093/pch/11.4.211] [Medline: 19030273]
110. Software as a medical device (SAMD): clinical evaluation : guidance for industry and Food and Drug Administration staff. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health. 2017 Dec 8. URL: https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101720008-pdf [accessed 2024-02-27]
111. Introduction to SAMD, IEC 62304 and IEC 82304-1. Medical Device HQ. URL: https://medicaldevicehq.com/introduction-to-samd-iec62304-and-iec-82304-1-online-course/ [accessed 2024-07-21]
112. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. BMC Med 2015 Jan 06;13:1 [FREE Full text] [doi: 10.1186/s12916-014-0241-z] [Medline: 25563062]
113. Ramakrishnan R, Rao S, He JR. Perinatal health predictors using artificial intelligence: a review. Womens Health (Lond) 2021;17:17455065211046132 [FREE Full text] [doi: 10.1177/17455065211046132] [Medline: 34519596]
114. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017 Jun 21;19:221-248 [FREE Full text] [doi: 10.1146/annurev-bioeng-071516-044442] [Medline: 28301734]


## Abbreviations

**AI:** artificial intelligence
**CDSS:** clinical decision support system
**ECG:** electrocardiogram
**EEG:** electroencephalography
**HR:** heart rate
**HRV:** heart rate variability
**MeSH:** Medical Subject Headings
**ML:** machine learning
**NICU:** neonatal intensive care unit
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**QMS:** quality management system
**SaMD:** software as a medical device
**SpO2:** oxygen saturation
**TRIPOD:** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis


*Edited by T de Azevedo Cardoso; submitted 02.03.23; peer-reviewed by I Provaznik, D Gopukumar, K Zhang, E Grooby; comments to author 16.01.24; revised version received 27.03.24; accepted 26.06.24; published 20.08.24.*

*Please cite as:*
*Rahman J, Brankovic A, Tracy M, Khanna S*
*Exploring Computational Techniques in Preprocessing Neonatal Physiological Signals for Detecting Adverse Outcomes: Scoping Review*
*Interact J Med Res 2024;13:e46946*
*URL: https://www.i-jmr.org/2024/1/e46946*
*doi:10.2196/46946*
*PMID:*

Review

# Parental Patterns of Alcohol Consumption During the COVID-19 Pandemic: Scoping Review

Christine Ou[1], MSN, PhD; Kathryn Corby[2], MScN; Kelsey Booth[2], BSN; Hui-Hui Ou[3], MA, MLIS

[1]Canadian Institute for Substance Use Research, School of Nursing, University of Victoria, Victoria, BC, Canada

[2]School of Nursing, University of Victoria, Victoria, BC, Canada

[3]Library Services, University of the Fraser Valley, Abbotsford, BC, Canada

**Corresponding Author:**
Christine Ou, MSN, PhD
Canadian Institute for Substance Use Research
School of Nursing
University of Victoria
HSD Building Room 402
3800 Finnerty Road
Victoria, BC, V8P 5C2
Canada
Phone: 1 250 721 7954
Email: christineou@uvic.ca

## Abstract

**Background:** The declaration of the COVID-19 pandemic led to public health restrictions that impacted the lives of people across the globe. Parents were particularly burdened with balancing multiple responsibilities, such as working from home while caring for and educating their children. Alcohol use among parents is an area that warrants further exploration.

**Objective:** This study aimed to investigate patterns of parental alcohol consumption during the COVID-19 pandemic, focusing on relative changes in the frequency and quantity of alcohol use compared to prepandemic use, nonparent adult samples, or both.

**Methods:** A scoping review informed by the methodology of Arksey and O'Malley explored patterns of parental alcohol consumption during the COVID-19 pandemic. Searches were conducted in CINAHL, Ovid MEDLINE, PsycINFO, and Web of Science. Search terms were created using the Joanna Briggs Institute framework of Population, Concept, and Context, with the population being parents and the concept being alcohol consumption during the COVID-19 pandemic.

**Results:** The database search yielded 3568 articles, which were screened for eligibility. Of the 3568 articles, 40 (1.12%) met the inclusion criteria and were included in the scoping review. Findings indicated the following: (1) having children at home was a factor associated with parental patterns of alcohol use; (2) mixed findings regarding gender-related patterns of alcohol consumption; and (3) linkages between parental patterns of alcohol use and mental health symptoms of stress, depression, and anxiety.

**Conclusions:** This scoping review revealed heterogeneous patterns in parental alcohol use across sociocultural contexts during the COVID-19 pandemic. Given the known harms of alcohol use, it is worthwhile for clinicians to assess parental drinking patterns and initiate conversations regarding moderation in alcohol use.

***(Interact J Med Res 2024;13:e48339)*** doi:10.2196/48339

**KEYWORDS**

parent; alcohol use; COVID-19; scoping review; parenting; alcohol; addict; addiction; substance use; health behavior; health behaviors; scoping; review methods; review methodology; drink; drinking; alcoholic; alcoholism

## Introduction

### Background

On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a global pandemic [1]. The restrictive public health measures introduced in many countries contributed to a shadow pandemic of psychological distress [2], which was associated with increased sales and consumption of alcohol [3,4]. Changes in the environments and circumstances in which adults drink can have effects on rates of consumption; lockdown

restrictions and sheltering in place led to drinking in the home environment becoming the norm during COVID-19 lockdowns in some places, such as the United Kingdom [5].

In Westernized countries, problematic alcohol use peaks in the third decade of life, a time when many adults are raising young children [6]. The reduction of alcohol consumption is one of the top 10 modifiable risk factors for reducing disease burden, injury, and social problems globally [7,8]. The burden of disease associated with alcohol use is high. A meta-analysis identified alcohol consumption as the seventh leading risk factor for disability and premature death in 2016, and among those aged 15 to 49 years, alcohol consumption accounts for nearly 10% of deaths on a global scale [9].

Parents warrant special attention as a large subsection of the adult population because they are primary caregivers for children. For parents, pandemic stressors (eg, lockdown restrictions, balancing employment while children are at home, and reduced social support) compound the daily stressors of parenting young children [10]. In parallel to the elevated rates of parental depression and anxiety from prepandemic levels [11,12], evidence suggests that the COVID-19 pandemic has increased the consumption of alcohol in parents with young children [13,14]. A meta-analysis of 128 studies (aggregate sample of N=492,235) revealed that nearly a quarter of adults reported increases in alcohol consumption during the COVID-19 pandemic [15]. These changes were moderated by per capita gross domestic product and country. The authors identified that residing with children was associated with increases in alcohol consumption, with consumption increasing with the number of children at home [15]. Moreover, while Acuff et al [15] identified that female participants were more likely to increase their drinking frequency and male participants were more likely to increase their problematic drinking behaviors (eg, binge drinking), it is unclear what proportions of these participants were also caregivers to young children. Currently, a granular and gendered examination of patterns of alcohol consumption in caregiving adults is lacking during the COVID-19 pandemic.

### Objectives

The purpose of this scoping review was to investigate broad patterns of parental alcohol consumption during the COVID-19 pandemic, examining the relative changes in the frequency and quantity of alcohol use compared to nonparent adults, prepandemic levels of consumption, or both. In addition to physical health harms related to excessive alcohol consumption [16], parents who consume problematic amounts of alcohol are more likely to have worse mental health and lower emotional availability to children [13]. Poor parental mental health, substance use, and negative parenting practices can have adverse consequences on children's socioemotional development and mental health [17,18]. The results of this scoping review can assist clinicians working with families in identifying parents at risk for alcohol misuse and engaging them in interventions to reduce consumption. The findings can also inform policy makers regarding the population of parents who may require targeted intervention and education on reducing and managing alcohol consumption in the aftermath of the COVID-19 pandemic.

## Methods

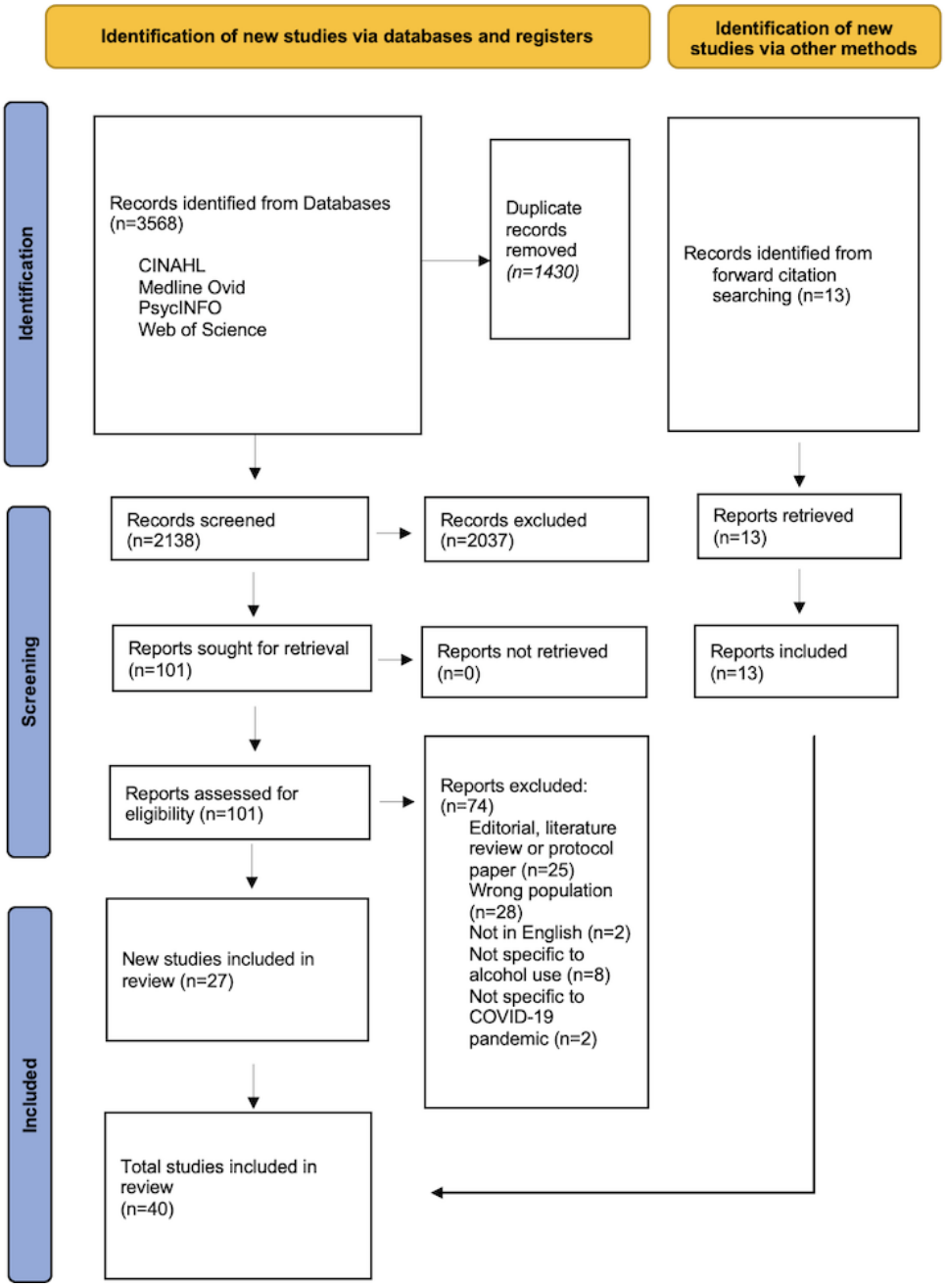
### Scoping Review Search Strategy

We undertook a scoping review following the methodology outlined by Arksey and O'Malley [19]. We used the Joanna Briggs Institute framework of Population, Concept, and Context (PCC) to create our search terms, with the population being parents (caregiving adults living with children aged <18 years in the same household) and the core concept being alcohol consumption within the context of the COVID-19 pandemic. To capture a wide scope of relevant studies, a librarian team member conducted a systematic search in CINAHL, Ovid MEDLINE, PsycINFO, Web of Science, and Cochrane databases using the specified PCC terms for all articles published after the WHO declaration of the COVID-19 pandemic on March 11, 2020 (Multimedia Appendix 1). The 3 groups of keywords within the PCC framework were joined with the Boolean operator "AND," which produced 3568 articles for screening in June 2022. Medical Subject Headings and its descriptors were used for CINAHL, Ovid MEDLINE, and PsycINFO. Forward citation searching of articles identified in June 2022 was carried out in May 2023, which further yielded an additional 13 articles.

### Inclusion and Exclusion Criteria

After 1430 (40.08%) duplicates were removed from the initial pool of 3568 articles, 2138 (59.92%) articles were screened for eligibility in a 2-round process (Figure 1). In the first round, article titles and abstracts (and text as needed) were reviewed against the PCC framework. We included peer-reviewed empirical studies that were published in English with data collection occurring after the WHO declaration of the COVID-19 pandemic. Nonempirical papers; expert opinions; letters to the editor; preprints; and empirical papers that were published after March 11, 2020, but did not contain alcohol consumption data collected during the COVID-19 pandemic were excluded. Studies that included a comparison of alcohol use between households with children and households without children were included. Reasons for exclusion included study population not being parents with alcohol consumption (eg, adolescent alcohol use) and the lack of explicit data collection on alcohol consumption as a variable. Studies that collected data on adult alcohol consumption during the COVID-19 pandemic but did not differentiate between adults living with children at home and adults not living with children at home were also excluded. Most studies that remained included both parent and nonparent participants (30/40, 75%), while some studies examined parents specifically (10/40, 25%).

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




### Data Extraction

Data from the articles were extracted into an extraction table derived from Polit and Beck [20]. The table included columns for authors, country of origin, year of publication, research design, sample size, measurement of alcohol use, time frame of data collection, and main findings. Extraction was performed independently by team members (KB and KC) and checked for accuracy (KC). Concerns with data extraction were resolved in consultation with the first author (CO).

### Evidence Synthesis

A narrative analysis of the main findings was conducted to identify and compare themes across the included studies. This approach allowed for the evaluation and integration of diverse findings on the trends in parental alcohol consumption during the COVID-19 pandemic.

## Results

### Overview

Of the 40 studies included in analyses, 36 (90%) were quantitative studies, 2 (5%) were qualitative studies, and 2 (5%) were systematic reviews (Figure 1). Of the 36 quantitative studies, most studies (n=28, 78%) were cross-sectional survey studies, 7 (19%) studies were longitudinal cohort studies, and 1 (3%) was a secondary data analysis of publicly available data (Table 1). The following themes were identified: (1) having children at home was a factor associated with parental patterns of alcohol use, (2) mixed patterns of alcohol consumption among mothers and fathers, and (3) heterogenous linkages between parental patterns of alcohol use and mental health.

**Table 1.** Evidence extraction table.

| Author and year | Time frame and country | Research design and sample size, N | Alcohol variable | Main findings |
|---|---|---|---|---|
| **Global** | | | | |
| Kyaw Hla et al [21], 2021 | • April 17 to June 25, 2020<br>• Global | • Cross-sectional survey<br>• N=37,206; n=28,649 (77%) women | • Respondents were asked about alcohol "stock up" and consumption frequency compared to pre–COVID-19 levels. | • Middle-aged, educated women working from home and living with children were a high-risk group for increased alcohol use during the lockdown. Households with children were more likely to increase alcohol use (ORa 1.17, 95% CI 1.10-1.24; P<.001) compared to households with adults only. |
| Roberts et al [4], 2021 | • December 2019 to November 2020<br>• Global | • Systematic review<br>• N=45 articles | • In the included studies, respondents were asked about patterns of alcohol use, harmful alcohol use, and binge drinking. | • Of the 45 studies, 5 (11%) studies found significantly higher alcohol use among parents compared to nonparents. |
| Sallie et al [22], 2020 | • May 12 to 28, 2020<br>• Global | • Cross-sectional survey<br>• N=1346; n=209 (15.53%) parents | • Respondents were asked about drinking behaviors before and during COVID-19 lockdowns.<br>• 3 items from the AUDIT-Cb | • Drinking behaviors decreased overall during quarantine across the entire sample. Among parents, drinking behaviors increased compared to nonparents (P=.003). |
| Schmidt et al [23], 2021 | • March 2020 to March 2021<br>• Data collection occurred between March and May 2020 in most articles (49/53, 92%).<br>• Global | • Systematic review<br>• N=53 articles | • Alcohol concept varied in the included studies. The PICOc tool was used to guide the search strategy where the problem identified was substance use, substance use disorder, drug abuse, and dependence. | • Caregiving responsibilities were discussed as a factor in increased substance use in 13 (25%) of the 53 articles. Parental status was associated with higher overall alcohol consumption, higher Alcohol Use Disorders Identification Test scores, consuming more drinks per occasion, more drinks consumed per week, and more heavy drinking episodes, with a positive correlation between the number of children at home and the amount of alcohol consumed. One study found that having children at home was associated with decreased alcohol use, and another study found that having children at home was associated with less binge drinking. |
| **North America** | | | | |
| Deacon et al [24], 2021 | • July 2020; reported retrospectively in April 2020<br>• Canada | • Cross-sectional survey<br>• N=758 couples; n=211 (27.8%) homeschooling couples; n=173 (22.8%) couples homeschooling due to the COVID-19 pandemic | • Brief Alcohol Motives Measure (2 items from the scale were used to examine coping-related alcohol use) | • Among the homeschooling sample, coping-related alcohol use was significantly increased relative to prepandemic use. The partner effect of hours spent homeschooling was significant on coping-related drinking. |

| Author and year | Time frame and country | Research design and sample size, N | Alcohol variable | Main findings |
|---|---|---|---|---|
| DesRoches et al [25], 2021 | • April to July 2020<br>• Canada | • Cross-sectional survey<br>• N=758 couples; n=211 (27.8%) homeschooling couples | • Quantity, frequency, and Peak Alcohol Use Index | • Women's hours of homeschooling were associated with greater drinking frequency in both the women themselves (beta=0.04; P=.01 and in their partner (beta=0.029; P=.02). Longer hours of homeschooling by men were associated with lower drinking frequency in women partners (b=–0.04; P=.02), but not with their own.<br>• Homeschooling was significantly correlated with drinking quantity (b=0.17; P<.05) but not drinking frequency. Partner's time spent home schooling was positively related to one's own drinking frequency (b=0.022; P=.01), quantity (b=0.020; P=.03), and peak drinking (b=0.022; P=.02). Gender significantly moderated the effect of time spent homeschooling on drinking frequency in dyadic analysis with both actor (P=.009) and partner (P<.001). |
| Gadermann et al [26], 2021 | • May 14 to 29, 2020<br>• Canada | • Cross-sectional survey<br>• N=3000; n=618 (20.6%) parents | • Respondents were asked about increased alcohol consumption after the pandemic as a means of coping with pandemic-related stress or deteriorating mental health.<br>• Questions were adapted from the Mental Health Foundation survey. | • Parents had a significantly greater increase in alcohol consumption compared to nonparents (28% vs 16%; P<.001). This was greater among men (32% vs 24%; P=.01). Parents were more likely to report deteriorated mental health (44.3% vs 35.6%; P<.001). |
| Hill MacEachern et al [27], 2021 | • September 11 to December 4, 2020<br>• Canada | • Cross-sectional survey<br>• N=12,344; n=3474 (28.14%) women and n=3348 (27.12%) men were parents or guardians to children aged <18 years | • Respondents were asked, "how has your alcohol consumption changed since before the COVID-19 pandemic?" | • Women with children were 1.46 times (95% CI 1.13-1.90) more likely to report increased alcohol consumption compared to women without children (23.3% vs 13.4%). Men with children were 1.38 times (95% CI 1.05-1.82) more likely to report increased consumption compared to men without children (21.7% vs 12.7%). |
| Joyce et al [28], 2022 | • April 14 to 28, 2020<br>• Canada | • Cross-sectional survey<br>• N=508 mothers of children aged 0 to 8 years | • Respondents were asked, "has your alcohol/drug use changed since the COVID-19 pandemic began?" | • Among the participants, 54.9% (n=279) reported no change in substance use, 39.2% (n=199) reported increased substance use, and 5.9% (n=30) reported decreased substance use. Alcohol was the most commonly reported substance (406/508, 80%), followed by cannabis (44/508, 8.7%). |
| Thomson et al [29], 2021 | • May 2020, September 2020, and January 2021<br>• Canada | • Multiround cross-sectional surveys<br>• May 2020: n=618 parents<br>• September 2020: n=804 parents<br>• January 2021: n=602 parents | • Respondents were asked whether their alcohol consumption had changed as a result of the COVID-19 pandemic. | • Parents were significantly more likely to report increased alcohol consumption compared to nonparents in all 3 rounds of surveys: round 1 (27.2% vs 16.1%), round 2 (21.9% vs 14.8%), and round 3 (22.4% vs 15.4%). Parents aged <35 years had higher odds of increased drinking than older parents (OR 1.51, 95% CI 1.10-2.07). Women were less likely than men to increase drinking (OR 0.76, 95% CI 0.59-0.97). Parents who continued to work while looking after children had higher odds of increased drinking (OR 1.86, 95% CI 1.45-2.40). |

| Author and year | Time frame and country | Research design and sample size, N | Alcohol variable | Main findings |
|---|---|---|---|---|
| Wardell et al [30], 2020 | • April to May 2020<br>• Canada | • Cross-sectional survey<br>• N=320; n=80 (25%) parents | • Respondents were asked about their alcohol consumption in the past 30 days and the month before the COVID-19 pandemic.<br>• Questions were modified from those provided by the National Institute on Alcohol Abuse and Alcoholism. | • Having a child in the household was a significant predictor of drinking as a coping behavior and drinking problems (β=.10; P<.05). |
| Zajacova et al [31], 2020 | • March 29 to April 3, 2020<br>• Canada | • Secondary data analysis<br>• N=4319 respondents; 30% (n=1296) had children aged <18 years | • Respondents were asked whether weekly alcohol consumption habits had changed. | • Those with children were more likely to report a decrease in alcohol consumption (P<.05). |
| Barbosa et al [32], 2023 | • February 2020, April 2020, July 2020, and November 2020<br>• United States | • Longitudinal survey design<br>• N=557; n=146 (26.3%) parents | • Respondents were asked about the quantity and frequency of alcohol consumption and binge drinking. | • The trajectory of alcohol consumption among parents found a 64% increase in the number of drinks consumed per month from February 2020 to November 2020. The increase in drinks consumed was significantly larger (P<.05) for those with children in the household than for those without children. |
| Boschuetz et al [33], 2020 | • April 5 to 12, 2020<br>• United States | • Cross-sectional survey<br>• N=408; 83% (338/405) were women and 80% (303/404) had children at home | • AUDIT-C | • Having children at home was associated with a significant increase in AUDIT-C scores (P=.02). |
| Freisthler and Price Wolf [34], 2023 | • April 2020, April 2021, and April 2022<br>• Ohio, United States | • 3 longitudinal waves of surveys, during April 2020, 2021, and 2022<br>• N=266 mothers across the 3 waves | • Mothers were asked how often they drank any kind of alcoholic beverage in the past year, on how many of the past 28 days they had at least 1 drink, and how many drinks were consumed in the past 28 days. | • Mothers reported fewer days of alcohol consumption in April 2021 and April 2022 compared to April 2020. However, the average number of drinks per day was higher in April 2021 and April 2022 than in April 2020.<br>• Mothers reported drinking less frequently but drinking more in volume when they did drink. |
| Grossman et al [35], 2020 | • May 2020<br>• United States | • Cross-sectional survey<br>• N=832; 45.1% (n=375) had children aged <18 years at home | • Respondents were asked about drinking frequency, binge drinking, patterns of drinking, and factors related to drinking in the past 30 days. | • Those with children at home consumed alcohol on a greater number of days than those without children (13.0 days vs 11.6 days; P=.054). There were no significant differences between parents and nonparents in the total drinks consumed or binge drinking. |
| Knell et al [36], 2020 | • April 15 to May 5, 2020<br>• United States | • Cross-sectional survey<br>• N=1804; n=785 (43.5%) parents | • BRFSSd<br>• If alcohol use was reported, further questions were asked about the “average number of daily drinks” and whether this number changed since the COVID-19 pandemic. | • Having children increased the odds of increased alcohol consumption (OR 1.58, 95% CI 1.19-2.09). |

| Author and year | Time frame and country | Research design and sample size, N | Alcohol variable | Main findings |
|---|---|---|---|---|
| Lamar et al [37], 2021 | • March 24 to April 28, 2020<br>• United States | • Cross-sectional survey<br>• N=1048 parents | • Alcohol Use Disorders Identification Test | • Problematic alcohol use was indicated among 26.5% (n=278) of the sample, with 11.4% (n=119) indicating a high level of problematic alcohol use. Male participants had significantly higher consumption compared to female participants (t1046=0.02; P=.003). |
| Nordeck et al [38], 2022 | • March to July 2020<br>• United States | • Longitudinal study<br>• 5 waves starting from March 11, 2020<br>• N=4298 across 5 surveys<br>• 29.3% (n=1259) living with children | • The number of drinking days in the past 7 days | • The number of drinking days was lower for participants living with a partner and children (β=–.65; 95% CI –0.82 to –0.48) and those living with children only (β=–.86; 95% CI –1.16 to –0.57) compared with participants without children. However, there were significant sustained increases in drinking days among those living with a partner and children compared to those living with children only or those in other household structures. |
| Pomazal et al [39], 2023 | • May 2020 to August 2021<br>• Wisconsin, United States | • Longitudinal study (3 waves)<br>• Wave 1: n=1290<br>• Wave 2: n=1868<br>• Wave 3: n=1585<br>• Percentage of participants with children at home at each wave: 29.4% (379/1290), 28.3% (528/1868), and 25.3% (401/1585), respectively. | • Individuals were asked to self-report alcohol consumption in the last 60 days (a lot more, a little more, same, little less, or much less) compared to the reference period (prepandemic period, then July 2020, and February 2021) | • In all 3 waves, the presence of children at home was associated with increased drinking (34.56%, 25.57%, and 22.38%; P<.001). Adjusted logistic regression model data: participants aged 55 years with children at home were less likely to increase drinking than those aged 35 years (wave 1: OR 0.23, 95% CI 0.1-0.53; wave 2: OR 0.4, 95% CI 0.17-0.91) and those aged 40 years (wave 1: OR 0.22, 95% CI 0.09-0.54; wave 2: OR 0.41, 95% CI 0.17-0.97) with children at home. |
| Rodriguez et al [40], 2021 | • July 22 to August 4, 2020<br>• United States | • Cross-sectional survey<br>• N=118 couples<br>• n=100 (84.75%) couples with children at home | • Respondents were asked about alcohol consumption patterns in the month before completing the survey.<br>• Items from Inventory of Problems–Alcohol and Drugs | • Having children at home was a significant predictor for drinking to cope (2.45; P<.05; CI 3.3-31.6). No significant association between having children at home and drinks consumed per week, high-intensity drinking, or alcohol-related problems. |
| Weerakoon et al [41], 2021 | • March to April 2020<br>• United States | • Cross-sectional survey<br>• N=1928; 42% (n=810) had children living in the household | • Respondents were asked about changes in drinking habits and binge drinking behaviors during the COVID-19 pandemic. | • Households with children were less likely to binge drink than those without children (AORe 0.74, 95% CI 0.58-0.94). No significant changes were reported in parental drinking compared with participants with no children living at home. |
| **Australia** | | | | |

| Author and year | Time frame and country | Research design and sample size, N | Alcohol variable | Main findings |
|---|---|---|---|---|
| Booth et al [42], 2024 | • September 2020<br>• Australia | • Cross-sectional survey<br>• N=4022 | • Respondents were asked to rank how often they drank alcohol before and during the lockdown on a scale from 1 (never) to 7 (≥2 times a day) | • Those with children were more likely to experience an increase in alcohol consumption (β=.51, 95% CI 0.37-0.76). |
| Callinan et al [43], 2021 | • April 29 to May 16, 2020<br>• Australia | • Cross-sectional survey<br>• N=2307; n=468 (20.3%) parents with dependent children | • Respondents were asked about their 2019 drinking behaviors (prepandemic period) and drinking behaviors in the past 30 days. | • Having dependent children was significantly associated with increased alcohol consumption (β=.62; 95% CI 0.32-0.92; P<.05), and homeschooling responsibilities were significantly associated with increased alcohol consumption (β=.53, 95% CI 0.2-0.82; P<.05). |
| Cook et al [13], 2021 | • July to September 2020<br>• Australia | • Qualitative study<br>• N=30 parents and caregivers of children aged 4 to 12 years | • Participants were asked about the nature of family lives before and after the COVID-19 pandemic, including changes in alcohol practices and family dynamics due to the COVID-19 pandemic. | • Alcohol use was reported to signal the end of the workday as a means of self-care and to alleviate boredom and manage stress. It was associated with feelings of guilt due to lockdown challenges. |
| Glenister et al [44], 2021 | • May 29 to July 9, 2020<br>• Australia | • Cross-sectional survey<br>• N=339 rural women; 41% (n=139) of sample lived with children | • Respondents were asked whether alcohol use increased, decreased, or remained the same since the COVID-19 pandemic. | • Rural women living with children were more likely to report increased alcohol consumption compared to rural women not living with children (OR 2.37, 95% CI 1.36-4.15; P=.002). |
| Greenwood et al [45], 2023 | • April 2020 to May 2021<br>• Victoria, Australia | • Longitudinal wave-based surveys (13 waves)<br>• N=2261 parents | • Parents were asked about the frequency of alcohol consumption per month (ranked on a 7-point scale). | • Estimated alcohol frequency trajectory for parents showed a decreased use over the course of the pandemic. Female and other parent gender were associated with a trajectory of lower frequency of alcohol use. Older parent age was associated with a trajectory of higher frequency of alcohol use. |
| Johnson et al [46], 2021 | • June to July 2020<br>• Australia | • Cross-sectional survey<br>• N=406 mothers | • Alcohol Use Disorders Identification Test<br>• Respondents were asked how their drinking has been impacted by the COVID-19 pandemic. | • Of the sample of 406 mothers, 54.9% (n=223) exceeded drinking guidelines and 41.4% (n=168) reported drinking more due to the pandemic. As parenting stress increased, alcohol use increased (P=.002). |
| Westrupp et al [47], 2023 | • April 8 to 28, 2020<br>• Australia | • Cross-sectional survey<br>• N=2365 parents | • Respondents were asked how often they drank alcoholic beverages. | • Compared to prepandemic population-based data from 4 Australian data sets, parents reported more frequent alcohol consumption (333/2365, 14.1%, pandemic data set) reported drinking on ≥4 days per week vs 771/9764, 7.89% (pre-pandemic data set), 7.9% reported drinking before the pandemic; P<.001).<br>• Women were less likely than mento consume alcohol at higher levels. |
| **Europe** | | | | |

| Author and year | Time frame and country | Research design and sample size, N | Alcohol variable | Main findings |
|---|---|---|---|---|
| Bramness et al [48], 2021 | • June to July 2020<br>• Norway | • Cross-sectional survey<br>• N=1328; of the n=1200 who reported any alcohol use, n=887 (66.79%) had children aged <18 years in the household | • Respondents were asked to report alcohol use in the past 12 months and changes in alcohol use since the COVID-19 measures were implemented.<br>• Two items from AUDIT-C | • Among the entire sample, 56.8% (n=754) reported no change in drinking patterns, 29.9% (n=397) reported less drinking, and 13.3% (n=177) reported more drinking. Having a child aged <18 years in the household was associated with more drinking (P=.02). |
| Koeger et al [49], 2022 | • April 2020 to January 2021<br>• Germany | • Multiround cross-sectional surveys<br>• Parents: round 1, n=307; round 2, n=295; and round 3, n=285 | • Participants were asked about AUFf on a weekly basis. | • Odds for an increased AUF was higher among participants with children in round 1 during the first lockdown (OR 1.34, 95% CI 0.92-1.96; P>.05) and in round 2 during the relaxation of lockdown restrictions (OR 1.77, 95% CI 1.118-2.65; P<.01). |
| Mangot-Sala et al [50], 2022 | • April 2020 to July 2021<br>• The Netherlands | • Longitudinal survey design<br>• N=63,194 | • Respondents were asked how many glasses of alcohol were consumed in the past 7 days. | • Results showed that during periods of lockdown, households with children reported the lowest alcohol consumption. During the summer, households with children reported a seasonal increase in drinking related to relaxed COVID-19 restrictions. |
| McAloney-Kocaman et al [51], 2022 | • March to June 2020<br>• United Kingdom | • Longitudinal study (multiwave survey)<br>• N=1268; 468 (36.9%) respondents had children in the household | • Participants were asked to indicate a change in alcohol consumption (drinking less, drinking approximately the same, or drinking more than usual) after the implementation of the lockdown. | • Perceived changes in alcohol consumption were significantly associated with the presence of children at home ($\chi$22=20.3, P<.001).<br>• Odds of increased alcohol consumption was lower for those with children in the household. |
| Oldham et al [52], 2021 | • April 30 to June 14, 2020<br>• United Kingdom | • Cross-sectional survey<br>• N=2777 parents and nonparents | • AUDIT-C | • Among men, living with children was significantly associated with increases in the units of alcohol consumed per drinking session (OR 1.72, 95% CI 1.09-2.73; P=.02) and the frequency of heavy episodic drinking (OR 2.40, 95% CI 1.44-3.99; P=.001). There was no significant increase in drinking for women living with children. |
| Thorell et al [53], 2022 | • April 28 to June 21, 2020<br>• Sweden, Spain, Italy, United Kingdom, Belgium, Netherlands, and Germany | • Cross-sectional survey<br>• N=6720 parents of children aged 5 to 18 years; n=5914 (88%) female | • Respondent were asked how their alcohol or drug changed use during the COVID-19 pandemic compared to their prepandemic use. | • Across the sample, 5% of homeschooling parents reported increased levels of alcohol or drug use compared to prepandemic use. Differences varied by country. In the United Kingdom, there was a 19.1% increase, whereas in Sweden, Spain, and Italy, <3% reported increased drinking or drug use. |
| Vanderbruggen et al [54], 2020 | • April 9 to 29, 2020<br>• Belgium | • Cross-sectional survey<br>• N=3632; 44.3% (n=1609) lived with children | • Respondents were asked about the average amount of alcohol consumed before and during the lockdown.<br>• Respondents were asked whether they drank more, less, or the same amount as before the COVID-19 pandemic. | • Overall, 30.3% (n=1100) of the total sample reported increased alcohol consumption, while 13.7% (n=498) reported decreased consumption. Alcohol use was positively correlated with the number of children living at home (22% increase in odds with every child at home; OR 1.220, 95% CI 1.146-1.289). |

| Author and year | Time frame and country | Research design and sample size, N | Alcohol variable | Main findings |
|---|---|---|---|---|
| Villette et al [55], 2022 | • January to March 2021<br>• Western Brittany, France | • Cross-sectional descriptive survey<br>• N=2491 | • AUDIT-C questionnaire used to assess change in alcohol consumption before, during, and after the lockdown (frequency of alcohol consumption, number of drinks per day, and frequency of heavy drinking) | • Of those living with family, 30.19% (468/1550; P<.001) experienced a greater increase in alcohol consumption than those living with adult roommates (21/135, 15.6%) and those living alone (89/395, 22.5%).<br>• Living with family was associated with increased alcohol consumption (OR 0.62, 95% CI 0.46-0.83; P<.001). |
| **Central and South America** | | | | |
| Garcia-Cerde et al, [56], 2021 | • May 22 to June 30, 2021<br>• Latin America | • Cross-sectional survey<br>• N=12,328; n=8136 (66%) female | • Respondents were asked to report on alcohol behaviors, including how often alcohol consumption occurred with children present. | • An overall decrease in drinking during the COVID-19 pandemic (77.5% vs 65%) was found, including drinking with children present.<br>• Regression model found that quarantining (β=3.81; 95% CI 2.61-5.02; P<.001), anxiety (β=.42; 95% CI 0.20-0.63; P<.001), and higher income were positively associated with drinking with children present. |
| **Asia** | | | | |
| Sugaya et al [57], 2021 | • June 15 to 20, 2021<br>• Japan | • Cross-sectional survey<br>• N=11,427; n=6388 (55.9%) parents | • Alcohol Use Disorders Identification Test (Japanese version) | • Those who answered “yes” to the “presence of child” had higher hazardous alcohol use (11.6% vs 9.5%), lower no-problem scores (81.2% vs 83.2%), and equal potential alcoholism scores (7.2%, respectively) compared to those who answered “no” ($\chi^2$=12.4; P=.002). |

[a]OR: odds ratio.

[b]AUDIT-C: Alcohol Use Disorders Identification Test–Consumption.

[c]PICO: Population, Intervention, Comparison, Outcome.

[d]BRFSS: Behavioral Risk Factor Surveillance System.

[e]AOR: adjusted odds ratio.

[f]AUF: alcohol use frequency.

Of the 40 included studies, 11 (28%) were also included in the 2 systematic reviews. To prevent undue inflation of support for the themes identified, we cited only the original papers to support specific themes.

## Having Children at Home as a Factor Associated With Parental Patterns of Alcohol Use

Many of the studies demonstrated linkages between parental alcohol use and the presence of children at home. Many studies found that, compared to not having children aged <18 years at home, having children at home was a significant predictor for an increase in alcohol consumption during the COVID-19 pandemic in the United States [32-36,38-40], Canada [26,27,29,30], Australia [42-44], Norway [48], Belgium [54], Germany [49], France [55], the United Kingdom [51,53], and Japan [57]. In a global survey study of 1346 adults (half of the sample were from the United Kingdom and United States), having children at home was a significant factor for increases in alcohol consumption, as operationalized by Alcohol Use Disorders Identification Test scores [22]. Participants in a nationally representative US sample of 2-parent households sustained increases in drinking days over the first 4 months of the COVID-19 pandemic, compared with the pre–COVID-19 baseline [38]. These findings are consistent with the conclusion of a higher incidence of increased drinking by adults with children at home compared with adults with no children at home in 2 systematic reviews during the COVID-19 pandemic [4,23]. Moreover, a large-scale global cross-sectional study involving 37,206 participants across 38 countries found households with children were significantly more likely to increase alcohol use compared to households with adults only [21]. An Australian survey of 4022 adults found that participants with children at home were more likely to increase the frequency of consuming alcohol along with other unhealthy foods (eg, snacks and sugared beverages) [42].

On the contrary, some studies indicated patterns of decreased parental drinking during the pandemic. In the United States, Weerakoon et al [41] found that there was no significant increase or decrease in parental drinking compared with participants with no children living at home (n=1928, 42% parents); moreover, having children in the household was associated with a decreased risk of binge drinking. Similarly, in Canada,

Zajacova et al [31] found that in a sample of 4319 adults (of whom, 30% (n=1296) had children aged <18 years at home), having children at home was associated with a lower rate of alcohol consumption. A 13-wave longitudinal study of Australian parents found that the frequency of alcohol use decreased over time, although there was no pre–COVID-19 consumption comparison [45]. In a large study of 35 Latin American countries (N=12,328), the majority of participants endorsed decreased drinking during the COVID-19 pandemic when compared with prepandemic levels of drinking [56].

Studies carried out in Europe indicate a mixture of findings. In a study of 63,194 adults in the Netherlands, Mangot-Sala et al [50] found that when comparing adults living with children at home to adults living without children at home and adults living alone, adults living with children drank less than the other 2 groups, suggesting that having children at home was protective in terms of drinking behaviors. Households with children reported a transient increase in drinking to prepandemic levels only during the summer months when restrictions were relaxed and families were more likely to gather, socialize, and engage in drinking with others [50]. In a European study of 6720 parents from (in the order of the largest number of participants per country) Germany, Sweden, Spain, Italy, Belgium, the United Kingdom, and the Netherlands, the proportion of parents who endorsed increased drinking during the COVID-19 pandemic (5%) were largely concentrated in the United Kingdom, with 19% (N=509) of UK parents reporting increases in drinking behaviors [53]. This contrasts with the findings of another UK-based longitudinal study, which found that having children at home was associated with lower odds of increased consumption [51]. A multiround cross-sectional study of adults in Germany found that participants who had children had higher odds of increased alcohol use frequency during the first lockdown (not significant) as well as during the easement of restrictions (significant) but lower odds during the second lockdown (not significant) [49].

In Australia, using a qualitative study of parents, Cook et al [13] found that while many parents in the sample indicated that they increased their frequency of alcohol consumption, some reported that they used the COVID-19 pandemic as an opportunity to lower their frequency of drinking through the absence of social opportunities for drinking.

## Mixed Patterns of Drinking Among Mothers and Fathers

Several studies indicated that fathers were more likely than mothers to increase drinking during the COVID-19 pandemic in North America [26,29,37], Australia [45,47], and the United Kingdom [52]. This is in contrast to a global study by Kyaw Hla et al [21], where they found middle-aged, educated women with children at home to be a high-risk group for increased alcohol use during lockdowns. Hill MacEachern et al [27] also identified that, among parent participants, women had slightly higher odds of reporting increases in alcohol consumption when compared with men. Homeschooling was a significant predictor for increased drinking during the pandemic in North America [24] and Australia [43]. Moreover, in a Canadian sample of parents, having to engage in homeschooling for children had an effect on maternal drinking. Desroches et al [25] found that mothers spent more time homeschooling than fathers and that both parents drank more when mothers spent more time homeschooling. Moreover, mothers drank less when fathers spent more time homeschooling [25]. Across their European sample (N=6720), Thorell et al [53] found that across subsamples from Sweden, Spain, Belgium, the Netherlands, Germany, Italy, and the United Kingdom, only 5% of homeschooling parents reported increased alcohol use. Freisthler and Price Wolf [34] investigated mothers' drinking patterns in a longitudinal study of US mothers via 3 waves of data collection (springtime of 2020, 2021, and 2022). Mothers reported significantly more days of alcohol consumption in the first wave when compared with the second and third waves; however, the average number of drinks consumed during a drinking day was greater in waves 2 and 3 [34]. In the qualitative study by Cook et al [13], one of the participants described how her own drinking habits changed while homeschooling during the pandemic:

> *But then by mid-April, I was completely out of work and we were still homeschooling, then drinking started about lunchtime, like come on, it's 12 o'clock, it's 5 o'clock somewhere, right.* [Woman, Queensland]

While Thomson et al [29] did not specifically examine homeschooling, they found that looking after children while working from home was associated with higher odds of increased alcohol consumption.

## Heterogenous Linkages Between Parental Patterns of Alcohol Use and Mental Health

Several studies looked at alcohol use as a coping mechanism for stress. In Australia, Johnson et al [46] found that parenting stress modestly correlated with the Alcohol Use Disorders Identification Test scores. In Canada, using path analysis, Wardell et al [30] found that having children at home was associated with greater alcohol consumption as a method of coping with pandemic stressors.

The findings regarding the relationships between alcohol consumption and mental health were mixed. Lamar et al [37] found a significant correlation between mental health symptoms (stress, depression, and anxiety) and increases in alcohol consumption in the United States. Qualitatively, parents in Australia described that partaking in alcohol consumption at the end of the day delineated a shift from time spent on the care of children to time for self and served a means of self-medicating stress and anxiety [13]. Garcia-Cerde et al [56] found anxiety to be a weak predictor of drinking with children present in Latin American countries. Thomson et al [29] and Joyce et al [28], in their Canadian studies, did not find a significant link between increased drinking and mental health symptoms, although Joyce et al [28] reported that mothers with a history of a previously existing anxiety disorder or elevated anxiety symptoms were more likely to increase their substance use. Notably, an alcohol use tracking app (Habit Tracker) that collected data from 83 countries (the majority from the United Kingdom and the United States) found that although participants with children (209/1134, 18.4% of the sample) indicated significant increases in drinking

severity based on their Alcohol Use Disorders Identification Test–Consumption scores, their levels of depression and anxiety were lower relative to adults with no children, suggesting a protective effect of having children at home [22].

## Discussion

### Principal Findings

The findings of this scoping review provide a broad examination of the patterns of parental alcohol consumption during the COVID-19 pandemic. The global patterns of parental alcohol use were heterogeneous and were influenced by the stage of COVID-19 data collection and sociocultural contexts. Most of the studies clustered around the first year and a half of the pandemic with regard to data collection. While lockdown measures were associated with increased frequency and quantity of alcohol consumption for some Western, industrialized countries, notably the United States, Canada, Australia, and the United Kingdom, a number of studies found patterns of decreased consumption. In Latin American and European Union countries especially, lockdown restrictions eliminated opportunities for socializing and drinking outside the home. Although these findings were mixed, there have been concerning reports of increased alcohol-related deaths; a recent Statistics Canada [58] report indicated an 18% increase in alcohol-related deaths during the COVID-19 pandemic when compared to previous years and that this was the largest change in alcohol-related deaths over the past 20 years. Similarly, in the United States, there were relative increases in deaths when comparing 2019 and 2020 rates of alcohol-related mortality [59], with childbearing-aged adults (ie, those aged 25 to 44 years) experiencing the largest increases. Angus et al [60] also found that the rate of alcohol-related deaths (characterized as "deaths of despair") increased in the United Kingdom during the COVID-19 pandemic. However, these reports did not distinguish adults who were parenting and those who were not.

Although we had expected to find an association between income support and alcohol use, many of the included studies did not focus on income as a predictor of alcohol use, with the exceptions of the study by Westrupp et al [47], which found a modest association between financial deprivation and lower alcohol use in Australia, and the study by Garcia-Cerde et al [56], which found that individuals with a higher income were more likely to drink with children present. McAloney-Kocaman et al [51], Greenwood et al [45], and Nordeck et al [38] did not separate parents from nonparents when examining the relationship between income and alcohol use and similarly found that lower income predicted decreased use. This contrasts with findings from the United Kingdom, which identified that the most socioeconomically disadvantaged households increased their alcohol purchases more than the least disadvantaged households [61].

In our study, we found some differences in the gender-related patterns of parental drinking, with some findings suggesting that fathers were more likely to increase overall alcohol consumption. Nonetheless, having to homeschool their children contributed to mothers' increases in drinking. Homeschooling and the provision of childcare during day care closures were responsibilities that fell disproportionately on mothers [62], with women also bearing greater employment-related consequences, such as the reduction of work hours and job loss related to childcare responsibilities [63]. The marketing of alcohol to mothers through social media was especially rampant (eg, through the use of hashtags such as #sendwine and slogans such as "from wifi to wine time #distance learning"), which calls into question the ethics of unmitigated marketing by the alcohol industry [10,14]. In the United States, gender-related differences in drinking behaviors have changed over time, with increases in rates of alcohol consumption in adult women [64]. Some studies indicate both mothers and fathers tend to decrease drinking during the transition to parenthood, with little difference between genders [65]. Meanwhile, other studies indicate that parenting children aged <1 year was associated with lower maternal drinking rates, whereas men's drinking habits changed little in response to parenthood [66]. Looking at overall drinking habits, fathers have been shown to be less likely to abstain from alcohol than mothers and consume greater volumes of alcohol [67]. Age has also been shown to be a factor, with young fathers more likely to partake in risky drinking behaviors than young mothers [67]. A systematic review of population-level alcohol policy interventions indicated evidence of gender-related differences in the impact of and exposure to alcohol marketing and failure to provide gender-specific recommendations [68].

Beyond the risk to parents' own health are the possible effects of parental alcohol consumption on children. Children are inevitably exposed to parental alcohol use in the home environment, and parents may even provide alcohol for older or adolescent children as a way to "safely" introduce children to alcohol in a parent-controlled home setting [69,70]. In the United States, Maggs et al [71] found that in 1 in 6 families, parents permitted their adolescents to drink in the home with the family during the COVID-19 pandemic, changing from prepandemic family practices of not permitting adolescent drinking at home. Evidence indicates that parents supplying even sips of alcohol to children carries increased risks of adverse alcohol outcomes, such as adolescent binge drinking, while parental supply of whole drinks was associated with higher odds of binge drinking, alcohol-related harms, and symptoms of dependence for teens [69]. Taken together, there is reason to continue to investigate parental patterns of alcohol consumption, given the range of negative short- and long-term health outcomes for children.

### Strengths and Limitations

Strengths of this scoping review include the systematic search process followed. The findings of this scoping review are limited by the heterogeneous ways in which alcohol consumption was measured and the fact that not all studies had data on patterns of consumption before the COVID-19 pandemic for comparison purposes. This review includes a large number of studies from early in the pandemic and hence does not provide the complete picture of the patterns of parental alcohol consumption from the beginning of the COVID-19 pandemic in 2020 to the removal of pandemic-related restrictions during 2022 and 2023. Moreover, although the included studies differentiated patterns of alcohol consumption among parenting adults and

nonparenting adults, a limited number of studies differentiated drinking habits between genders in parenting adults, which limited the provision of a more granular examination of the association between gender and alcohol consumption habits in parents.

### Clinical Implications and Future Directions

Parental drinking can adversely affect children indirectly through parent preoccupation and the diversion of parental attention and supervision and directly through physical and verbal violence in the home setting and influence children's later drinking habits [67,72]. In the 2023 *Canada's Guidance on Alcohol and Health* [73] report, the Canadian Centre for Substance Use and Addiction indicated that consuming >2 standard drinks per week heightens risks for breast and colon cancers, heart disease, and stroke, with the consumption of >6 drinks per week representing high risk of harms. In line with these strong recommendations for reducing alcohol intake to decrease health and social risks, clinicians who work with families (eg, primary care providers, nurses, and social workers) can inquire about clients' frequency and amount of alcohol consumption and counsel strategies for moderating or reducing consumption. The findings of this review suggest that parents caring for children are a population that requires more empirical investigation in relation to the amounts and frequency of alcohol consumption and problematic alcohol use. Gender–based data analysis of parental drinking behaviors is also important for informing interventions and policies to promote safe alcohol use among parents. Because parents effectively serve as role models for children, their drinking habits can influence children's later drinking behaviors. It is important to better understand how to assist parents in moderating alcohol intake, given the risks of unmitigated alcohol consumption.

### Conclusions

Our scoping review indicated that the COVID-19 pandemic influenced patterns of parental alcohol consumption in different ways. Sociocultural influences contributed to determining whether having children at home was a protective or risk factor for alcohol consumption. In countries where drinking alcohol is more likely to occur in social settings, such as at bars or restaurants (eg, Latin American and some European countries), parental consumption tended to decrease during the lockdown restrictions, while in countries where drinking in the home environment is the norm (eg, the United States, Canada, the United Kingdom, and Australia), parental consumption tended to increase. Given the known harms of alcohol, clinicians can initiate conversations about parental drinking habits and counsel moderation for parents who report amounts of drinking higher than national guidelines.

## Acknowledgments

The authors are grateful for Hana Kim's assistance with forward citation searching to identify additional relevant studies that cited key articles in our scoping review.

This scoping review was funded by the University of Victoria School of Nursing Dorothy Kergin Endowment Award.

## Data Availability

Full data extraction matrix is available upon request.

## Authors' Contributions

CO contributed to study conceptualization, funding acquisition, methodology, project administration, supervision, formal analysis, and writing the original draft. KC contributed to investigation, formal analysis, and writing the original draft. KB contributed to formal analysis and writing the original draft. HHO contributed to data curation and reviewing and editing the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 37 KB - ijmr_v13i1e48339_app1.docx ]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 502 KB - ijmr_v13i1e48339_app2.pdf ]

## References

1. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
2. Zaidi A, Ali AZ. Living under the shadow of a pandemic: the psychological challenges underlying social distancing and awareness raising. Psychol Trauma 2020 Jul;12(5):508-510. [doi: 10.1037/tra0000815] [Medline: 32525374]

3. Lee BP, Dodge JL, Leventhal A, Terrault NA. Retail alcohol and tobacco sales during COVID-19. Ann Intern Med 2021 Jul;174(7):1027-1029 [FREE Full text] [doi: 10.7326/M20-7271] [Medline: 33646843]
4. Roberts A, Rogers J, Mason R, Siriwardena AN, Hogue T, Whitley GA, et al. Alcohol and other substance use during the COVID-19 pandemic: a systematic review. Drug Alcohol Depend 2021 Dec 01;229(Pt A):109150 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.109150] [Medline: 34749198]
5. Conroy D, Nicholls E. 'When I open it, I have to drink it all': push and pull factors shaping domestic alcohol consumption during the COVID-19 pandemic UK Spring 2020 lockdown. Drug Alcohol Rev 2022 Sep;41(6):1275-1283 [FREE Full text] [doi: 10.1111/dar.13396] [Medline: 34708473]
6. Delker E, Brown Q, Hasin D. Alcohol consumption in demographic subpopulations: an epidemiologic overview. Alcohol Res 2016;38(1):7-15 [FREE Full text] [Medline: 27159807]
7. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
8. Vergés A, Jackson KM, Bucholz KK, Grant JD, Trull TJ, Wood PK, et al. Deconstructing the age-prevalence curve of alcohol dependence: why "maturing out" is only a small piece of the puzzle. J Abnorm Psychol 2012 May;121(2):511-523 [FREE Full text] [doi: 10.1037/a0026027] [Medline: 22060948]
9. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018 Sep 22;392(10152):1015-1035 [FREE Full text] [doi: 10.1016/S0140-6736(18)31310-2] [Medline: 30146330]
10. Adams EL, Smith D, Caccavale LJ, Bean MK. Parents are stressed! Patterns of parent stress across COVID-19. Front Psychiatry 2021 Apr 8;12:626456 [FREE Full text] [doi: 10.3389/fpsyt.2021.626456] [Medline: 33897489]
11. Feinberg ME, A Mogle J, Lee JK, Tornello SL, Hostetler ML, Cifelli JA, et al. Impact of the COVID-19 pandemic on parent, child, and family functioning. Fam Process 2022 Mar;61(1):361-374 [FREE Full text] [doi: 10.1111/famp.12649] [Medline: 33830510]
12. Racine N, Hetherington E, McArthur BA, McDonald S, Edwards S, Tough S, et al. Maternal depressive and anxiety symptoms before and during the COVID-19 pandemic in Canada: a longitudinal analysis. Lancet Psychiatry 2021 May;8(5):405-415 [FREE Full text] [doi: 10.1016/S2215-0366(21)00074-2] [Medline: 33773109]
13. Cook M, Dwyer R, Kuntsche S, Callinan S, Pennay A. 'I'm not managing it; it's managing me': a qualitative investigation of Australian parents' and carers' alcohol consumption during the COVID-19 pandemic. Drugs Educ Prev Policy 2021 Jul 15;29(3):308-316. [doi: 10.1080/09687637.2021.1950125]
14. Basch CH, Meleo-Erwin ZC, Mohlman J, Fera J, Quinones N. Use of the Instagram hashtags #winemom and #momjuice among mothers during the COVID-19 pandemic: descriptive, cross-sectional study. JMIR Pediatr Parent 2021 May 18;4(2):e28991 [FREE Full text] [doi: 10.2196/28991] [Medline: 33848257]
15. Acuff SF, Strickland JC, Tucker JA, Murphy JG. Changes in alcohol use during COVID-19 and associations with contextual and individual difference variables: a systematic review and meta-analysis. Psychol Addict Behav 2022 Feb;36(1):1-19 [FREE Full text] [doi: 10.1037/adb0000796] [Medline: 34807630]
16. Rehm J, Gmel GES, Gmel G, Hasan OS, Imtiaz S, Popova S, et al. The relationship between different dimensions of alcohol use and the burden of disease-an update. Addiction 2017 Jun;112(6):968-1001 [FREE Full text] [doi: 10.1111/add.13757] [Medline: 28220587]
17. Uddin J, Alharbi N, Uddin H, Hossain MB, Hatipoğlu SS, Long DL, et al. Parenting stress and family resilience affect the association of adverse childhood experiences with children's mental health and attention-deficit/hyperactivity disorder. J Affect Disord 2020 Jul 01;272:104-109. [doi: 10.1016/j.jad.2020.03.132] [Medline: 32379600]
18. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The adverse childhood experiences (ACE) study. Am J Prev Med 1998 May;14(4):245-258. [doi: 10.1016/s0749-3797(98)00017-8] [Medline: 9635069]
19. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
20. Polit DF, Beck CT. Nursing Research: Generating and Assessing Evidence for Nursing Practice, Eighth Edition. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
21. Kyaw Hla Z, Ramalho R, Teunissen L, Cuykx I, Decorte P, Pabian S, et al. Socioeconomic and environmental factors associated with increased alcohol purchase and consumption in 38 countries during the Covid-19 pandemic. Front Psychiatry 2021 Jan 14;12:802037 [FREE Full text] [doi: 10.3389/fpsyt.2021.802037] [Medline: 35095616]
22. Sallie SN, Ritou V, Bowden-Jones H, Voon V. Assessing international alcohol consumption patterns during isolation from the COVID-19 pandemic using an online survey: highlighting negative emotionality mechanisms. BMJ Open 2020 Nov 26;10(11):e044276 [FREE Full text] [doi: 10.1136/bmjopen-2020-044276] [Medline: 33243820]
23. Schmidt RA, Genois R, Jin J, Vigo D, Rehm J, Rush B. The early impact of COVID-19 on the incidence, prevalence, and severity of alcohol use and other drugs: a systematic review. Drug Alcohol Depend 2021 Nov 01;228:109065 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.109065] [Medline: 34600257]

24. Deacon SH, Rodriguez LM, Elgendi M, King FE, Nogueira-Arjona R, Sherry SB, et al. Parenting through a pandemic: mental health and substance use consequences of mandated homeschooling. Couple Fam Psychol Res Pract 2021 Dec;10(4):281-293. [doi: 10.1037/cfp0000171]
25. DesRoches DI, Deacon SH, Rodriguez LM, Sherry SB, Nogueira-Arjona R, Elgendi MM, et al. Homeschooling during COVID-19: gender differences in work–family conflict and alcohol use behaviour among romantic couples. Soc Sci 2021 Jun 23;10(7):240. [doi: 10.3390/socsci10070240]
26. Gadermann AC, Thomson KC, Richardson CG, Gagné M, McAuliffe C, Hirani S, et al. Examining the impacts of the COVID-19 pandemic on family mental health in Canada: findings from a national cross-sectional study. BMJ Open 2021 Jan 12;11(1):e042871 [FREE Full text] [doi: 10.1136/bmjopen-2020-042871] [Medline: 33436472]
27. Hill MacEachern K, Venugopal J, Varin M, Weeks M, Hussain N, Baker MM. Applying a gendered lens to understanding self-reported changes in alcohol and cannabis consumption during the second wave of the COVID-19 pandemic in Canada, September to December 2020. Health Promot Chronic Dis Prev Can 2021 Nov 10;41(11):331-339 [FREE Full text] [doi: 10.24095/hpcdp.41.11.03] [Medline: 34569771]
28. Joyce KM, Cameron EE, Sulymka J, Protudjer JL, Roos LE. Changes in maternal substance use during the COVID-19 pandemic. J Stud Alcohol Drugs 2022 May;83(3):352-357. [Medline: 35590175]
29. Thomson KC, Jenkins E, Gill R, Richardson CG, Gagné Petteni M, McAuliffe C, et al. Impacts of the COVID-19 pandemic on family mental health in Canada: findings from a multi-round cross-sectional study. Int J Environ Res Public Health 2021 Nov 17;18(22):12080 [FREE Full text] [doi: 10.3390/ijerph182212080] [Medline: 34831830]
30. Wardell JD, Kempe T, Rapinda KK, Single A, Bilevicius E, Frohlich JR, et al. Drinking to cope during COVID-19 pandemic: the role of external and internal factors in coping motive pathways to alcohol use, solitary drinking, and alcohol problems. Alcohol Clin Exp Res 2020 Oct;44(10):2073-2083. [doi: 10.1111/acer.14425] [Medline: 32870516]
31. Zajacova A, Jehn A, Stackhouse M, Denice P, Ramos H. Changes in health behaviours during early COVID-19 and socio-demographic disparities: a cross-sectional analysis. Can J Public Health 2020 Dec;111(6):953-962 [FREE Full text] [doi: 10.17269/s41997-020-00434-y] [Medline: 33170494]
32. Barbosa C, Dowd WN, Barnosky A, Karriker-Jaffe KJ. Alcohol consumption during the first year of the COVID-19 pandemic in the United States: results from a nationally representative longitudinal survey. J Addict Med 2023;17(1):e11-e17 [FREE Full text] [doi: 10.1097/ADM.0000000000001018] [Medline: 35861345]
33. Boschuetz N, Cheng S, Mei L, Loy VM. Changes in alcohol use patterns in the United States during COVID-19 pandemic. WMJ 2020 Sep;119(3):171-176 [FREE Full text] [Medline: 33091284]
34. Freisthler B, Price Wolf J. Longitudinal changes in self-reported alcohol consumption by mothers' during the COVID-19 pandemic. Alcohol Alcohol 2023 May 09;58(3):235-237 [FREE Full text] [doi: 10.1093/alcalc/agac064] [Medline: 36521494]
35. Grossman ER, Benjamin-Neelon SE, Sonnenschein S. Alcohol consumption during the COVID-19 pandemic: a cross-sectional survey of US adults. Int J Environ Res Public Health 2020 Dec 09;17(24):9189 [FREE Full text] [doi: 10.3390/ijerph17249189] [Medline: 33316978]
36. Knell G, Robertson MC, Dooley EE, Burford K, Mendez KS. Health behavior changes during COVID-19 pandemic and subsequent "stay-at-home" orders. Int J Environ Res Public Health 2020 Aug 28;17(17):6268 [FREE Full text] [doi: 10.3390/ijerph17176268] [Medline: 32872179]
37. Lamar MR, Speciale M, Forbes LK, Donovan C. The mental health of U.S. parents during the COVID-19 pandemic. J Ment Health Couns 2021;43(4):319-335. [doi: 10.17744/mehc.43.4.03]
38. Nordeck CD, Riehm KE, Smail EJ, Holingue C, Kane JC, Johnson RM, et al. Changes in drinking days among United States adults during the COVID-19 pandemic. Addiction 2022 Feb 12;117(2):331-340 [FREE Full text] [doi: 10.1111/add.15622] [Medline: 34159674]
39. Pomazal R, Malecki KM, McCulley L, Stafford N, Schowalter M, Schultz A. Changes in alcohol consumption during the COVID-19 pandemic: evidence from Wisconsin. Int J Environ Res Public Health 2023 Mar 29;20(7):5301 [FREE Full text] [doi: 10.3390/ijerph20075301] [Medline: 37047917]
40. Rodriguez LM, Litt DM, Stewart SM. COVID-19 psychological and financial stress and their links to drinking: a dyadic analysis in romantic couples. Psychol Addict Behav 2021 Jun;35(4):377-390. [doi: 10.1037/adb0000724] [Medline: 34096760]
41. Weerakoon SM, Jetelina KK, Knell G. Longer time spent at home during COVID-19 pandemic is associated with binge drinking among US adults. Am J Drug Alcohol Abuse 2021 Jan 02;47(1):98-106. [doi: 10.1080/00952990.2020.1832508] [Medline: 33280423]
42. Booth L, Santos JA, Hoek AC, Webster J, Pettigrew S. Population subgroups at risk of unhealthy changes in food and beverage consumption during COVID-19 lockdowns. Int J Behav Med 2024 Feb 06;31(1):163-168 [FREE Full text] [doi: 10.1007/s12529-023-10165-2] [Medline: 36879051]
43. Callinan S, Mojica-Perez Y, Wright CJ, Livingston M, Kuntsche S, Laslett AM, et al. Purchasing, consumption, demographic and socioeconomic variables associated with shifts in alcohol consumption during the COVID-19 pandemic. Drug Alcohol Rev 2021 Feb 10;40(2):183-191. [doi: 10.1111/dar.13200] [Medline: 33170976]

44. Glenister KM, Ervin K, Podubinski T. Detrimental health behaviour changes among females living in rural areas during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Jan 15;18(2):722 [FREE Full text] [doi: 10.3390/ijerph18020722] [Medline: 33467693]
45. Greenwood CJ, Fuller-Tyszkiewicz M, Hutchinson DM, Macdonald JA, Bereznicki HG, Youssef GJ, et al. Alcohol use among Australian parents during the COVID-19 pandemic - April-2020 to May 2021. Addict Behav 2023 Mar;138:107561 [FREE Full text] [doi: 10.1016/j.addbeh.2022.107561] [Medline: 36473249]
46. Johnson E, Fellowes R, Cant K, Hunt S. Self-assured and sober: the relationship between maternal parenting sense of competence, stress, and alcohol use. Front Glob Womens Health 2021 Jan 31;2:778183 [FREE Full text] [doi: 10.3389/fgwh.2021.778183] [Medline: 35174356]
47. Westrupp EM, Bennett C, Berkowitz T, Youssef GJ, Toumbourou JW, Tucker R, et al. Child, parent, and family mental health and functioning in Australia during COVID-19: comparison to pre-pandemic data. Eur Child Adolesc Psychiatry 2023 Feb;32(2):317-330 [FREE Full text] [doi: 10.1007/s00787-021-01861-z] [Medline: 34417875]
48. Bramness JG, Bye EK, Moan IS, Rossow I. Alcohol use during the COVID-19 pandemic: self-reported changes and motives for change. Eur Addict Res 2021 Apr 9;27(4):257-262 [FREE Full text] [doi: 10.1159/000515102] [Medline: 33839730]
49. Koeger M, Schillok H, Voss S, Coenen M, Merkel C, Jung-Sievers C. Alcohol use of German adults during different pandemic phases: repeated cross-sectional analyses in the COVID-19 snapshot monitoring study (COSMO). Int J Environ Res Public Health 2022 May 01;19(9):5489 [FREE Full text] [doi: 10.3390/ijerph19095489] [Medline: 35564883]
50. Mangot-Sala L, Tran KA, Smidt N, Liefbroer AC. The impact of the COVID lockdown on alcohol consumption in the Netherlands. The role of living arrangements and social isolation. Drug Alcohol Depend 2022 Apr 01;233:109349 [FREE Full text] [doi: 10.1016/j.drugalcdep.2022.109349] [Medline: 35168117]
51. McAloney-Kocaman K, McPherson KE, McGlinchey E, Armour C. Factors associated with changing alcohol consumption during the first UK lockdown. Eur J Public Health 2022 Oct 03;32(5):766-772 [FREE Full text] [doi: 10.1093/eurpub/ckac124] [Medline: 36094148]
52. Oldham M, Garnett C, Brown J, Kale D, Shahab L, Herbec A. Characterising the patterns of and factors associated with increased alcohol consumption since COVID-19 in a UK sample. Drug Alcohol Rev 2021 Sep 03;40(6):890-899 [FREE Full text] [doi: 10.1111/dar.13256] [Medline: 33657644]
53. Thorell LB, Skoglund C, de la Peña AG, Baeyens D, Fuermaier AB, Groom MJ, et al. Parental experiences of homeschooling during the COVID-19 pandemic: differences between seven European countries and between children with and without mental health conditions. Eur Child Adolesc Psychiatry 2022 Apr;31(4):649-661 [FREE Full text] [doi: 10.1007/s00787-020-01706-1] [Medline: 33415470]
54. Vanderbruggen N, Matthys F, van Laere S, Zeeuws D, Santermans L, van den Ameele S, et al. Self-reported alcohol, tobacco, and cannabis use during COVID-19 lockdown measures: results from a web-based survey. Eur Addict Res 2020;26(6):309-315 [FREE Full text] [doi: 10.1159/000510822] [Medline: 32961535]
55. Villette PA, Lyonnard O, Trehu C, Barais M, Le Goff D, Le Floch B, et al. Changes in alcohol consumption after 1 year of the COVID-19 pandemic: a cross-sectional study in a region of France. Int J Environ Res Public Health 2022 Nov 15;19(22):15049 [FREE Full text] [doi: 10.3390/ijerph192215049] [Medline: 36429768]
56. Garcia-Cerde R, Valente JY, Sohi I, Falade R, Sanchez ZM, Monteiro MG. Alcohol use during the COVID-19 pandemic in Latin America and the Caribbean. Rev Panam Salud Publica 2021 May 20;45:e52 [FREE Full text] [doi: 10.26633/RPSP.2021.52] [Medline: 34025727]
57. Sugaya N, Yamamoto T, Suzuki N, Uchiumi C. Alcohol use and its related psychosocial effects during the prolonged COVID-19 pandemic in Japan: a cross-sectional survey. Int J Environ Res Public Health 2021 Dec 17;18(24):13318 [FREE Full text] [doi: 10.3390/ijerph182413318] [Medline: 34948939]
58. Provisional death counts and excess mortality, January 2020 to October 2022. Statistics Canada. 2023 Jan 12. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/230112/dq230112c-eng.htm [accessed 2023-04-25]
59. White AM, Castle IJ, Powell PA, Hingson RW, Koob GF. Alcohol-related deaths during the COVID-19 pandemic. JAMA 2022 May 03;327(17):1704-1706 [FREE Full text] [doi: 10.1001/jama.2022.4308] [Medline: 35302593]
60. Angus C, Buckley C, Tilstra AM, Dowd JB. Increases in 'deaths of despair' during the COVID-19 pandemic in the United States and the United Kingdom. Public Health 2023 May;218:92-96 [FREE Full text] [doi: 10.1016/j.puhe.2023.02.019] [Medline: 36996743]
61. Anderson P, O'Donnell A, Jané Llopis E, Kaner E. The COVID-19 alcohol paradox: British household purchases during 2020 compared with 2015-2019. PLoS One 2022 Jan 19;17(1):e0261609 [FREE Full text] [doi: 10.1371/journal.pone.0261609] [Medline: 35045099]
62. O'Sullivan K, Rock N, Burke L, Boyle N, Joksimovic N, Foley H, et al. Gender differences in the psychosocial functioning of parents during the COVID-19 pandemic. Front Psychol 2022 Jul 8;13:846238 [FREE Full text] [doi: 10.3389/fpsyg.2022.846238] [Medline: 35899000]
63. Petts RJ, Carlson DL, Pepin JR. A gendered pandemic: childcare, homeschooling, and parents' employment during COVID‐19. Gender Work Org 2021 Jan 13;28(S2):515-534. [doi: 10.1111/gwao.12614]

64. Keyes KM, Jager J, Mal-Sarkar T, Patrick ME, Rutherford C, Hasin D. Is there a recent epidemic of women's drinking? A critical review of national studies. Alcohol Clin Exp Res 2019 Jul 05;43(7):1344-1359 [FREE Full text] [doi: 10.1111/acer.14082] [Medline: 31074877]
65. Evans-Polce RJ, Jang BJ, Maggs JL, Patrick ME. Gender and age differences in the associations between family social roles and excessive alcohol use. Soc Sci Med 2020 Jan;244:112664 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112664] [Medline: 31726267]
66. Borschmann R, Becker D, Spry E, Youssef GJ, Olsson CA, Hutchinson DM, et al. Alcohol and parenthood: an integrative analysis of the effects of transition to parenthood in three Australasian cohorts. Drug Alcohol Depend 2019 Apr 01;197:326-334 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.02.004] [Medline: 30878883]
67. Laslett AM, Jiang H, Room R. Alcohol consumption of Australian parents: continuity and change in the new millennium. Foundation for Alcohol Research and Education. 2017. URL: https://fare.org.au/wp-content/uploads/Alcohol-consumption-of-Australian-parents-FINAL-30-JUNE-2017.pdf [accessed 2024-07-24]
68. Fitzgerald N, Angus K, Emslie C, Shipton D, Bauld L. Gender differences in the impact of population-level alcohol policy interventions: evidence synthesis of systematic reviews. Addiction 2016 Oct;111(10):1735-1747 [FREE Full text] [doi: 10.1111/add.13452] [Medline: 27177685]
69. Aiken A, Clare PJ, Boland VC, Degenhardt L, Yuen WS, Hutchinson D, et al. Parental supply of sips and whole drinks of alcohol to adolescents and associations with binge drinking and alcohol-related harms: a prospective cohort study. Drug Alcohol Depend 2020 Oct 01;215:108204. [doi: 10.1016/j.drugalcdep.2020.108204] [Medline: 32871506]
70. Sigman A. Covid-19 and alcohol: parental drinking influences the next generation. BMJ 2020 Jun 25;369:m2525. [doi: 10.1136/bmj.m2525] [Medline: 32586844]
71. Maggs JL, Cassinat JR, Kelly BC, Mustillo SA, Whiteman SD. Parents who first allowed adolescents to drink alcohol in a family context during spring 2020 COVID-19 emergency shutdowns. J Adolesc Health 2021 Apr;68(4):816-818 [FREE Full text] [doi: 10.1016/j.jadohealth.2021.01.010] [Medline: 33582017]
72. Rossow I, Felix L, Keating P, McCambridge J. Parental drinking and adverse outcomes in children: a scoping review of cohort studies. Drug Alcohol Rev 2016 Jul;35(4):397-405 [FREE Full text] [doi: 10.1111/dar.12319] [Medline: 26332090]
73. Canada's guidance on alcohol and health. Canadian Centre on Substance Use and Addiction. URL: https://www.ccsa.ca/canadas-guidance-alcohol-and-health [accessed 2023-04-26]


## Abbreviations

**AUDIT:** Alcohol Use Disorders Identification Test
**PCC:** Population, Concept, and Context
**WHO:** World Health Organization


*Edited by T de Azevedo Cardoso; submitted 01.06.23; peer-reviewed by J Egunjobi; comments to author 19.06.24; revised version received 21.06.24; accepted 24.06.24; published 26.08.24.*

*Please cite as:*
*Ou C, Corby K, Booth K, Ou HH*
*Parental Patterns of Alcohol Consumption During the COVID-19 Pandemic: Scoping Review*
*Interact J Med Res 2024;13:e48339*
*URL: https://www.i-jmr.org/2024/1/e48339*
*doi:10.2196/48339*
*PMID:*

Review

# Central Hemodynamic and Thermoregulatory Responses to Food Intake as Potential Biomarkers for Eating Detection: Systematic Review

Lucy Chikwetu[1], BA, MSc, PhD; Parker Vakili[2], BSc; Andrew Takais[2], BSc; Rabih Younes[1], BE, MSE, PhD

[1]Department of Electrical and Computer Engineering, Duke University, Durham, NC, United States

[2]Department of Computer Science, Duke University, Durham, NC, United States

**Corresponding Author:**
Rabih Younes, BE, MSE, PhD
Department of Electrical and Computer Engineering
Duke University
210 Hudson Hall
Durham, NC, 27708
United States
Phone: 1 919 660 5051
Email: rabih.younes@duke.edu

## Abstract

**Background:** Diet-related diseases, such as type 2 diabetes, require strict dietary management to slow down disease progression and call for innovative management strategies. Conventional diet monitoring places a significant memory burden on patients, who may not accurately remember details of their meals and thus frequently falls short in preventing disease progression. Recent advances in sensor and computational technologies have sparked interest in developing eating detection platforms.

**Objective:** This review investigates central hemodynamic and thermoregulatory responses as potential biomarkers for eating detection.

**Methods:** We searched peer-reviewed literature indexed in PubMed, Web of Science, and Scopus on June 20, 2022, with no date limits. We also conducted manual searches in the same databases until April 21, 2024. We included English-language papers demonstrating the impact of eating on central hemodynamics and thermoregulation in healthy individuals. To evaluate the overall study quality and assess the risk of bias, we designed a customized tool inspired by the Cochrane assessment framework. This tool has 4 categories: high, medium, low, and very low. A total of 2 independent reviewers conducted title and abstract screening, full-text review, and study quality and risk of bias analysis. In instances of disagreement between the 2 reviewers, a third reviewer served as an adjudicator.

**Results:** Our search retrieved 11,450 studies, and 25 met our inclusion criteria. Among the 25 included studies, 32% (8/25) were classified as high quality, 52% (13/25) as medium quality, and 16% (4/25) as low quality. Furthermore, we found no evidence of publication bias in any of the included studies. A consistent postprandial increase in heart rate, cardiac output, and stroke volume was observed in at least 95% (heart rate: 19/19, cardiac output: 18/19, stroke volume: 11/11) of the studies that investigated these variables' responses to eating. Specifically, cardiac output increased by 9%-100%, stroke volume by 18%-41%, and heart rate by 6%-21% across these studies. These changes were statistically significant ($P$<.05). In contrast, the 8 studies that investigated postprandial thermoregulatory effects displayed grossly inconsistent results, showing wide variations in response with no clear patterns of change, indicating a high degree of variability among these studies.

**Conclusions:** Our findings demonstrate that central hemodynamic responses, particularly heart rate, hold promise for wearable-based eating detection, as cardiac output and stroke volume cannot be measured by any currently available noninvasive medical or consumer-grade wearables.

**Trial Registration:** PROSPERO CRD42022360600; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=360600

***(Interact J Med Res 2024;13:e52167)*** doi:10.2196/52167

**KEYWORDS**

eating detection; eating moment recognition; postprandial physiological responses; postprandial physiology; eating; food; consumption; postprandial; hemodynamics prandial; thermoregulation; physiological; heart rate; vital; vitals; wearable; wearables; thermoregulatory hemodynamic; biomarker; biomarkers; diet; dietary; monitoring; detect; detection; detecting; synthesis; review methods; review methodology; systematic; sensor; sensors; digital health

## Introduction

The rising incidence of diet-related diseases, such as coronary heart disease [1], and type 2 diabetes [2] has led to the emergence of an innovative research field called automated diet monitoring [3,4]. The primary objective of this field is to advance technologies that facilitate comprehensive monitoring of critical elements of food intake, such as meal timing, duration, quantity, and nutritional composition [5]. A fundamental aspect of this field is eating detection [4], which involves using technologies such as wearable devices to determine when an individual is eating. This innovative approach holds immense potential in empowering individuals to more accurately and efficiently monitor their dietary habits and effectively manage chronic diseases that require precise dietary control. By leveraging wearable technologies, eating detection enables real-time monitoring of meal timing, duration, and eating patterns, thus providing valuable insights into eating behaviors and supporting overall health and well-being.

Despite extensive research in eating detection [6-10], successful deployment of eating detection platforms remains elusive. Several challenges have hindered their widespread adoption, including the use of custom-made wearables with impractical form factors [4,11] or privacy concerns [12]. Notable examples include dental implants [11], on-body cameras [12], and wrist-worn inertial sensors [13,14] (accelerometer, gyroscope, or magnetometer), which pose significant challenges for widespread acceptance. In addition, many studies have relied on data collected in controlled laboratory or semicontrolled field settings [6,15,16], leading to algorithms that struggle to perform effectively in real-life scenarios filled with diverse activities and situations.

Furthermore, while some studies have achieved accurate detection of eating episodes [17-19], few have demonstrated the feasibility of real-time implementation [6], raising questions about their suitability for use in free-living scenarios. Real-time computation requires high-performance computing platforms with long battery life, presenting significant challenges for many proposed solutions [6]. In addition, detecting eating episodes during concurrent activities, such as walking, remains a substantial hurdle that necessitates further algorithmic improvements [7].

By the same token, most of the existing systems rely on a single sensor with limited provisions for sensor failure or suboptimal performance. In contrast, multimodal sensing, which involves using different sensors on one or more devices, can provide complementary and unique information, enhancing the performance and reliability of eating detection platforms, especially when combined with higher sampling rates [4,20].

Given the aforementioned challenges, extensive research focusing on exploring untapped sensing signals and repurposing everyday wearable devices for eating detection remains crucial. In line with this objective, our review investigates the potential of central hemodynamic and thermoregulatory responses to food intake as promising biomarkers for eating detection. Through a comprehensive analysis of the existing literature, this review aims to shed light on the role of central hemodynamics and thermoregulation in monitoring eating behavior, contributing valuable insights toward the development of effective and practical eating detection platforms.

## Methods

### Overview

This review adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [21] guidelines. It is registered on PROSPERO (CRD42022360600).

### Information Sources

We searched peer-reviewed literature indexed in PubMed, Web of Science, and Scopus on June 20, 2022, with no date limits. We also conducted manual searches in the same databases until April 21, 2024. Before the official search, we identified a key study [22] and decided to augment our search by retrieving officially provided similar studies from PubMed and Web of Science.

### Search Strategy

This review examines central hemodynamic and thermoregulatory responses to food intake as potential biomarkers for wearable-based eating detection. To capture relevant studies, our search used keywords such as “hemodynamic,” “haemodynamic*,” “thermoregulat*,” “temperature regulation,” “body heat,” “body temperature,” “skin temperature,” “eating,” “meal*,” “ingest*,” “intak*,” “postprandial,” and “post-prandial.” We excluded all animal studies from our search. While our search strategy did not impose language restrictions on retrieved papers, we excluded 10 non–English-language papers published in Czech, French, German, and Japanese. For a comprehensive overview of our search strategy across all databases, please refer to Multimedia Appendix 1.

### Eligibility Criteria

All included studies were peer-reviewed journal and conference papers published before April 21, 2024. Eligible studies had to demonstrate the impact of eating on central hemodynamics and thermoregulation, assessing metrics such as changes in heart rate, blood pressure, cardiac output, core or skin temperature, or stroke volume in healthy individuals.

We excluded non-English papers, animal research studies, opinion pieces, letters, studies not specifically focusing on central hemodynamics (eg, those focusing on renal dynamics), studies using indirect calorimetry to measure postprandial

thermogenesis, studies testing drug effects, studies focusing on water ingestion, long-term effects of food, and studies primarily examining central hemodynamic and thermoregulatory responses to food intake during or after exercise.

### Screening and Selection

All retrieved studies were exported to Covidence [23], which automatically identified and removed 1419 duplicates. In total, 2 reviewers independently conducted the title and abstract screening and the full-text review. In cases of disagreement between the 2 reviewers, a third reviewer acted as an adjudicator.

### Data Extraction and Synthesis

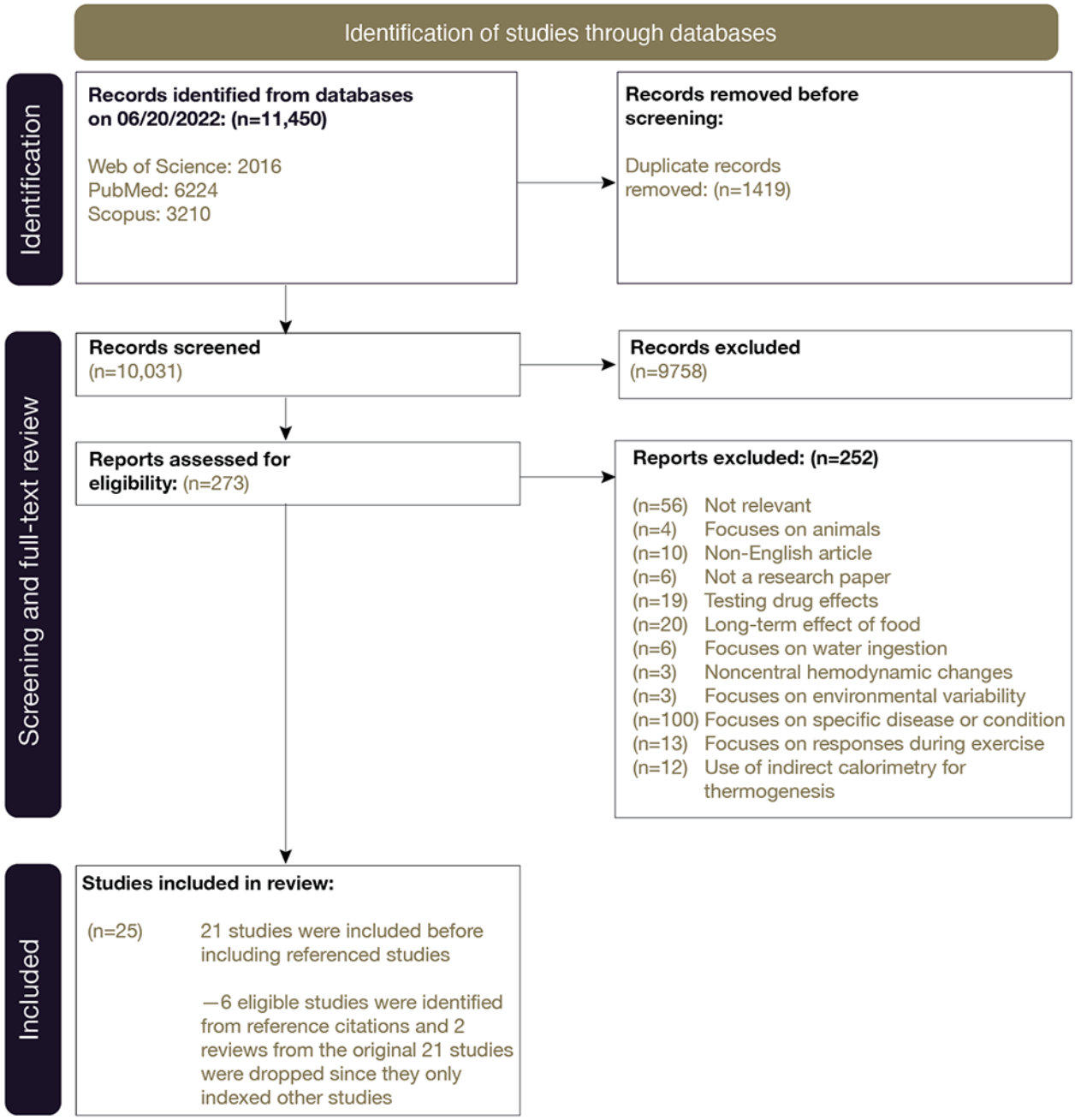
A total of 2 reviewers independently conducted data extraction, study quality assessment, and risk-of-bias assessment for each included study. Any conflicts or discrepancies were resolved by an adjudicator. In addition, if any of the included studies referenced other studies that met our eligibility criteria, we also extracted relevant information from those referenced studies. We extracted 8 study characteristics from the included studies (Multimedia Appendix 2).

Due to the heterogeneity in the outcomes and designs of the included studies, we used a narrative data synthesis approach. In our characteristic tables (Multimedia Appendix 2), we summarized key findings from each study, identifying common themes, patterns, and differences. This approach allowed us to integrate both quantitative and qualitative data, thus providing a comprehensive overview of the current landscape. Finally, all graphs were generated using ggplot2 in R (version 4.0.2; R Foundation for Statistical Computing).

## Results

Our search retrieved 11,450 studies (1419 duplicates), resulting in 10,031 studies being screened (Figure 1). Following the title and abstract screening, 273 studies progressed to the full-text review. Out of these, 21 studies met the eligibility criteria. In addition, we conducted a nested search within the references of the 21 studies and identified 6 additional relevant studies. We excluded 2 studies from the original 21 as they were review papers referencing other studies already included in our analysis. Finally, we extracted data from the included 25 studies and proceeded to analyze them.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram illustrating the article selection process.




To assess overall study quality and risk of bias, we developed a custom tool comprising 4 categories: high, medium, low, and very low (Multimedia Appendix 3). This assessment tool was based on the Cochrane assessment framework [24]. In total, 2 reviewers independently evaluated the study quality and risk of bias for each study, with an adjudicator resolving any conflicts. Among the 25 included studies, 32% (8/25) were classified as high quality, 52% (13/25) as medium quality, and 16% (4/25) as low quality (Multimedia Appendix 3). Furthermore, we found no evidence of publication bias in any of the included studies.

Of the included studies, 14 physiological responses were recorded. Multimedia Appendix 4 presents a breakdown of the number of studies reporting each postprandial physiological response. The number of participants per study ranged from 4 to 104, with a mean of 17 (SD 19). Out of the 416 participants in the 25 included studies, 230 (55.3%) were male, 180 (44.3%) were female, and the remaining 6 (1.4%) had unknown sex. The age of the participants in the included studies ranged from 18 to 69 years, and all participants were healthy. The sessions varied in duration, spanning from 10 minutes to 8 hours, and the provided food items included cake, cheese, filet mignon, and boiled eggs, among other foods.

Our findings overwhelmingly demonstrate a significant increase in heart rate after eating; 19 included studies investigated heart rate, and all 19 studies showed a statistically significant ($P<.05$) rise ranging from 6% to 21% (Multimedia Appendix 2). Furthermore, the postprandial effects of heart rate were found to be generally similar in both supine and erect positions (Multimedia Appendix 2). In addition, there was strong evidence

supporting an increase in cardiac output after food ingestion, with 18 of the 19 studies on cardiac output showing a statistically significant ($P$<.05) postprandial increase ranging from 9% to 100%, and only 1 study showing a statistically insignificant ($P$>.05) response for a high-fat liquid meal consisting of emulsified peanut oil (Multimedia Appendix 2). Similarly, data from 11 studies revealed strong evidence supporting a postprandial increase in stroke volume, ranging from 18% to 41% (Multimedia Appendix 2). These results collectively highlight the consistent and significant cardiovascular responses associated with food intake.

Out of the 25 included studies, 16 (64%) investigated the response of blood pressure to food ingestion; however, the results were inconclusive (Multimedia Appendix 2). Among these studies, 7 observed a postprandial increase in systolic blood pressure, while 4 found a statistically insignificant ($P$>.05) response, and 1 study even reported a postprandial decrease. Similarly, 8 studies found a postprandial decrease in diastolic blood pressure, while 3 studies found a statistically insignificant ($P$>.05) response, and 1 study observed a postprandial increase. Regarding mean blood pressure, the findings indicated either inconsistent postprandial behavior or a statistically insignificant ($P$>.05) response (Multimedia Appendix 2).

A total of 2 studies investigated the response of vascular resistance to food ingestion; one study focused solely on systemic vascular resistance, while the other study researched both systemic and mesenteric vascular resistance (Multimedia Appendix 2). Both studies observed a statistically significant ($P$<.05) postprandial decrease in vascular resistance.

In addition, 6 studies examined the response of blood flow in the hand, calf, and superior mesenteric artery to food ingestion (Multimedia Appendix 2). However, the postprandial effects of blood flow in the hand or calf were unclear. A total of 2 studies indicated a statistically significant ($P$<.05) postprandial increase in hand and calf blood flow, while 3 studies found a statistically insignificant ($P$>.05) change (Multimedia Appendix 2). Furthermore, 5 studies explored the response of oxygen uptake to food intake. In total, 3 of these studies observed a postprandial increase in oxygen uptake (more details in Multimedia Appendix 2). In addition, 1 study showed an inconsistent response, and another reported a statistically insignificant change ($P$>.05). Finally, the postprandial effects of skin or core temperature were found to be grossly inconsistent (Multimedia Appendix 2). A total of 5 studies reported inconsistent responses of skin or core temperature to food ingestion, 2 studies observed a postprandial increase, and 1 study found a statistically insignificant ($P$>.05) response (Multimedia Appendix 2).

## Discussion

The primary objective of this systematic review was to investigate central hemodynamic and thermoregulatory responses to food intake as potential biomarkers for detecting eating events. Among the 25 studies included in this review, at least 95% (heart rate: 19/19, cardiac output: 18/19, stroke volume: 11/11) of those investigating the response of heart rate, cardiac output, and stroke volume to food ingestion reported consistent, statistically significant ($P$<.05) elevations in these variables. In contrast, the postprandial thermoregulatory effects were markedly inconsistent across the 8 studies that investigated them.

Our findings provide valuable insights into the physiological changes that occur after food consumption and shed light on potential biomarkers for detecting eating events. They consistently demonstrate significant cardiovascular responses associated with food intake (Multimedia Appendix 2). Specifically, heart rate showed a significant increase after eating, as evidenced by 19 studies, with the rise ranging from 6% to 21% (Multimedia Appendix 4 and Textbox 1). Similarly, cardiac output and stroke volume exhibited a robust postprandial increase, with 18 studies on cardiac output showing a significant rise ranging from 11% to 100%, and 11 studies revealing a postprandial increase in stroke volume, ranging from 18% to 41% (Multimedia Appendix 4 and Textbox 1). These findings suggest that heart rate, cardiac output, and stroke volume may serve as reliable biomarkers for detecting eating events. However, it is worth mentioning that 1 study reported a statistically insignificant ($P$>.05) cardiac output response to a high-fat liquid meal [25], while 9 studies reported different postprandial effects depending on the composition of the consumed food, underscoring the potential influence of dietary composition on cardiovascular reactions. Consequently, it is worth exploring whether eating detection using cardiovascular responses, particularly heart rate, could be enhanced to sense macronutrients in the ingested food. A particular area of focus could be carbohydrate-aware eating detection platforms, which would have the ability to classify the amount of carbohydrates in the food a person has just eaten into 2 classes that are high-carbohydrate and low-carbohydrate. We anticipate this to be feasible since several studies have demonstrated that postprandial heart rate changes within the first hour of eating correlate with the carbohydrate content in the food an individual has just consumed [25-28]. Accordingly, individuals could receive alerts when their meals are carbohydrate-rich, thus empowering them with more information about the foods they consume. In addition, decision support systems for diabetes management could use the inferred postprandial carbohydrate information to improve patient outcomes by personalizing patient recommendations based on their diet information.

**Textbox 1.** Postprandial percentage increase range for cardiac output, stroke volume, and heart rate.

**Postprandial percentage increase range**

- Cardiac output: 9%-100%
- Stroke volume: 18%-41%
- Heart rate: 6%-21%

Currently, there are no noninvasive consumer or medical-grade wearables capable of measuring cardiac output and stroke volume. As a result, heart rate remains the only signal that could feasibly be used in everyday eating detection platforms. Studies using heart rate for eating detection are already underway, but they have primarily focused on animal studies so far [29,30]. A notable study conducted in humans used consumer smartwatches and successfully detected eating events using heart rate with an accuracy of 98.6% [31]. However, the sensitivity, specificity, and $F_1$-scores were very low, with an $F_1$-score as low as 2% in one of the experiments [31]. We anticipate that using high-resolution data and conducting initial experiments in stationary situations without vigorous physical activity could facilitate a better understanding of heart rate–based eating detection systems.

Conversely, the findings concerning blood pressure responses were inconclusive. Of the 25 studies included, 16 (64%) investigated blood pressure changes following food ingestion; however, the results were inconsistent for systolic, diastolic, and mean blood pressure changes (Multimedia Appendix 2). The lack of consensus on blood pressure responses may limit its potential as a reliable biomarker for eating detection using wearable devices.

Vascular resistance and blood flow were also examined in a subset of the included studies (n=2 and n=6, respectively). Interestingly, vascular resistance showed a consistent postprandial decrease in both studies that investigated this parameter (Multimedia Appendix 2). On the other hand, the effects of postprandial blood flow in the hand, calf, and superior mesenteric artery were inconclusive, with some studies indicating postprandial effects, while others found no significant changes (Multimedia Appendix 2). Consequently, the use of vascular resistance and blood flow as biomarkers for eating detection may require further investigation and validation.

Furthermore, oxygen uptake responses after food intake were explored in 5 studies (Multimedia Appendix 2). While 3 studies showed a postprandial increase in oxygen uptake, 1 study reported inconsistent results, and another found no significant change. These mixed findings suggest that oxygen uptake may have limited use as a standalone biomarker for wearable-based eating detection.

Finally, postprandial temperature effects were found to be grossly inconsistent (Multimedia Appendix 2). While many studies have focused on determining the thermic effect of food [32-34] through calorimetry, very little is known about the skin or core temperature response to eating, and this review aimed to bridge that gap. Among the included studies, 2 reported a postprandial increase in temperature, while 5 studies revealed inconsistent responses, and 1 study found no significant change (Multimedia Appendix 2). The inconsistency in temperature responses warrants caution when considering temperature-based biomarkers for eating detection. In addition, temperature can be easily affected by environmental factors, and there is often a time lag before individuals elicit any temperature change in response to any physiological or environmental factors.

While our systematic review provides valuable insights, there are several limitations that should be acknowledged. First, the available literature on central hemodynamic and thermoregulatory responses to food intake might be limited, potentially leading to a restricted pool of studies to learn from. Sample sizes and study methodologies varied significantly, and this heterogeneity could have affected the ability to draw definitive conclusions for many variables. Furthermore, the majority of the studies focused on heart rate, cardiac output, and stroke volume, leaving other potential biomarkers underexplored. The inconsistency in the postprandial effects on blood pressure, blood flow, oxygen uptake, and temperature further highlights the need for more comprehensive research in these areas. In addition, it is important to note that all included studies were based on healthy participants, so the results from this review might not be fully extensible to the broader population, and their applicability to individuals with specific health conditions remains uncertain.

Out of the 25 included studies, 19 (76%) were conducted before 2000 using contemporary devices, which might have compromised the results' quality compared with more recent studies that may have benefitted from the advances in modern technology. Recent advancements in precision, data collection, and analytical tools could significantly improve the accuracy and reliability of these findings. All studies included in this review had the goal of wanting to understand the postprandial effects of the concerned signals, and they were measuring changes in a discrete fashion as opposed to a continuous fashion. Furthermore, the quality assessment of the included studies indicated that 17 (68%) of the studies were categorized as medium or low quality, mostly because they did not provide information on the devices used or the frequency at which data was collected. While efforts were made to minimize bias during the review process, the quality of data reported in the primary studies could influence the overall robustness of our findings.

Overall, the insights gained from this systematic review provide a strong foundation for future research and the development of wearable technologies aimed at enhancing eating detection and monitoring. Central hemodynamic responses, particularly heart rate, offer promising prospects for wearable-based eating detection. As other identified biomarkers, such as cardiac output and stroke volume cannot be measured by any currently available noninvasive medical or consumer-grade wearable, heart rate remains a valuable signal for this purpose.

Given the consistent postprandial heart rate effects in both supine and erect positions, heart rate–based eating detection technologies have the potential to revolutionize assisted living care, particularly in enhancing monitoring and support for bedridden individuals. For instance, this technology could track and verify that patients are receiving meals at prescribed times, providing reassurance to doctors and family members that dietary guidelines are being followed and proper care is being administered. Despite its potential, the complexity of the heart rate signal, influenced by factors such as physical exercise [35] and stress responses [36], may complicate its ability to detect eating events. To accurately identify eating episodes, algorithms must account for these variables and differentiate between heart rate changes due to eating and other activities. Integrating heart

rate data with additional sensors, such as electrodermal activity (EDA) sensors, which measure skin conductance and emotional states [37,38], and inertial measurement units (IMUs) sensors, such as accelerometers and gyroscopes, which monitor physical activity or hand movements, could enhance algorithm performance. This multimodal approach increases context awareness, allowing for clearer distinctions between physiological changes from eating and other activities, such as exercise or emotional states. By reducing false positives, such systems can become more sensitive and specific to changes associated with eating. It is also important to note, however, that many signals, including heart rate, EDA, and IMUs, carry the risk of reidentification [39]. Therefore, any systems using these signals should be properly secured to mitigate this privacy concern.

In conclusion, central hemodynamic responses, particularly heart rate, show promise for wearable-based eating detection. Future studies could aim for larger sample sizes, standardized protocols, and well-controlled experimental conditions to enhance the generalizability and reliability of the results. In addition, investigating how these findings apply to diverse populations and individuals with specific health conditions is crucial for broader application and impact.

## Acknowledgments

The authors would like to express their gratitude to Sarah Park, the engineering and computer science librarian at Duke University, for her invaluable assistance in designing the search strategy. No funding was received for this study.

## Data Availability

All data generated in this study are available in the appendices. The search strategy can be found in Multimedia Appendix 1, and the list of extracted study characteristics for each included study is in Multimedia Appendix 2. In addition, the risk-of-bias and study quality assessments are provided in Multimedia Appendix 3, and the postprandial physiological responses, along with the number of studies reporting data for each category, are in Multimedia Appendix 4.

## Authors' Contributions

LC conceived the study. LC, AT, and PV designed the search strategy. LC retrieved all searches from the databases and exported them to Covidence. LC, AT, and PV conducted the title and abstract screening, full-text review, and data extraction for this study. LC created all tables, graphs, and other visual representations, and AT and PV participated in the quality assurance process for all tables, graphs, and visual representations. LC drafted the manuscript, and all other authors read, edited, and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 38 KB - ijmr_v13i1e52167_app1.pdf ]

Multimedia Appendix 2
Characteristics of included studies.
[PDF File (Adobe PDF File), 196 KB - ijmr_v13i1e52167_app2.pdf ]

Multimedia Appendix 3
Study quality and publication bias assessment.
[PDF File (Adobe PDF File), 164 KB - ijmr_v13i1e52167_app3.pdf ]

Multimedia Appendix 4
Postprandial physiological responses and the number of studies reporting data for each category.
[PDF File (Adobe PDF File), 28 KB - ijmr_v13i1e52167_app4.pdf ]

Multimedia Appendix 5
PRISMA Checklist.
[DOCX File , 30 KB - ijmr_v13i1e52167_app5.docx ]

## References

1. Vaduganathan M, Mensah GA, Turco JV, Fuster V, Roth GA. The global burden of cardiovascular diseases and risk: a compass for future health. J Am Coll Cardiol 2022;80(25):2361-2371 [FREE Full text] [doi: 10.1016/j.jacc.2022.11.005] [Medline: 36368511]
2. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes - global burden of disease and forecasted trends. J Epidemiol Glob Health 2020;10(1):107-111 [FREE Full text] [doi: 10.2991/jegh.k.191028.001] [Medline: 32175717]
3. Hassannejad H, Matrella G, Ciampolini P, De Munari I, Mordonini M, Cagnoni S. Automatic diet monitoring: a review of computer vision and wearable sensor-based methods. Int J Food Sci Nutr 2017;68(6):656-670. [doi: 10.1080/09637486.2017.1283683] [Medline: 28139173]
4. Prioleau T, Moore Li E, Ghovanloo M. Unobtrusive and wearable systems for automatic dietary monitoring. IEEE Trans Biomed Eng 2017;64(9):2075-2089. [doi: 10.1109/TBME.2016.2631246] [Medline: 28092510]
5. Chikwetu L, Daily S, Mortazavi BJ, Dunn J. Automated diet capture using voice alerts and speech recognition on smartphones: pilot usability and acceptability study. JMIR Form Res 2023;7:e46659 [FREE Full text] [doi: 10.2196/46659] [Medline: 37191989]
6. Nyamukuru M, Odame K. Tiny eats: eating detection on a microcontroller. 2020 Presented at: Proceedings of 2020 IEEE Second Workshop on Machine Learning on Edge in Sensor Systems (SenSys-ML); April 21, 2020; Sydney, NSW, Australia p. 19-23. [doi: 10.1109/sensysml50931.2020.00011]
7. Sharma S, Jasper P, Muth E, Hoover A. The impact of walking and resting on wrist motion for automated detection of meals. ACM Trans. Comput. Healthcare 2020;1(4):1-19. [doi: 10.1145/3407623]
8. Olubanjo T, Ghovanloo M. Real-time swallowing detection based on tracheal acoustics. 2014 Presented at: Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 4-9, 2014; Florence, Italy p. 4384-4388. [doi: 10.1109/icassp.2014.6854430]
9. Saphala A, Zhang R, Amft O. Proximity-based eating event detection in smart eyeglasses with expert and data models. New York, NY, United States: ACM; 2022 Presented at: Proceedings of the 2022 ACM Int Symp Wearable Comput; September 11-15, 2022; Cambridge, United Kingdom p. 59-63. [doi: 10.1145/3544794.3558476]
10. Gao Y, Zhang N, Wang H, Ding X, Ye X, Chen G, et al. iHear food: eating detection using commodity bluetooth headsets. Piscataway, New Jersy: IEEE; 2016 Presented at: Proceedings of the 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); June 27-29, 2016; Washington, DC, USA URL: https://www.researchgate.net/publication/306301557_iHear_Food_Eating_Detection_Using_Commodity_Bluetooth_Headsets [doi: 10.1109/chase.2016.14]
11. Li CY, Chen YC, Chen WJ, Huang P, Chu H. Sensor-embedded teeth for oral activity recognition. New York, NY, United States: ACM Press; 2013 Presented at: Proceedings of the 17th Annual International Symposium on Wearable Computers - ISWC 13; September 8-12, 2013; Zurich, Switzerland. [doi: 10.1145/2493988.2494352]
12. Thomaz E, Parnami A, Essa I, Abowd G. Feasibility of identifying eating moments from first-person images leveraging human computation. New York, NY, US: ACM Press; 2013 Presented at: Proceedings of the 4th Int SenseCam Pervasive Imaging Conference - SenseCam '13; November 18-19, 2013; San Diego, California. [doi: 10.1145/2526667.2526672]
13. Sharma S, Jasper P, Muth E, Hoover A. The impact of walking and resting on wrist motion for automated detection of meals. ACM Trans. Comput. Healthcare 2020;1(4):1-19. [doi: 10.1145/3407623]
14. Zhang S, Alharbi R, Nicholson M, Alshurafa N. When generalized eating detection machine learning models fail in the field. New York, NY, US: ACM; 2017 Presented at: Proceedings of the 2017 ACM Int Jt Conf Pervasive Ubiquitous Comput Proc 2017 ACM Int Symp Wearable Comput; September 11-17, 2017; Maui, Hawaii. [doi: 10.1145/3123024.3124409]
15. Sazonov E, Schuckers S, Lopez-Meyer P, Makeyev O, Sazonova N, Melanson EL, et al. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. Physiol Meas 2008;29(5):525-541 [FREE Full text] [doi: 10.1088/0967-3334/29/5/001] [Medline: 18427161]
16. Kandori A, Yamamoto T, Sano Y, Oonuma M, Miyashita T, Murata M, et al. Simple magnetic swallowing detection system. IEEE Sensors J 2012;12(4):805-811. [doi: 10.1109/jsen.2011.2166954]
17. Walker WP, Bhatia DK. Automated ingestion detection for a health monitoring system. IEEE J Biomed Health Inform 2014;18(2):682-692. [doi: 10.1109/JBHI.2013.2279193] [Medline: 24043408]
18. Xu Y, Guanling C, Yu C. Automatic eating detection using head-mount and wrist-worn accelerometers. Piscataway, New Jersey: IEEE; 2015 Presented at: Proceedings of the 2015 17th Int Conf E-Health Netw Appl Serv Heal; October 14-17, 2015; Boston, MA p. 578. [doi: 10.1109/healthcom.2015.7454568]
19. Hotta S, Mori T, Uchida D, Maeda K, Yaginuma Y, Inomata A. Eating moment recognition using heart rate responses. New York, NY, United States: ACM; 2017 Presented at: Proceedings of the 2017 ACM Int Jt Conf Pervasive Ubiquitous Comput Proc 2017 ACM Int Symp Wearable Comput; September 11-15, 2017; Maui, Hawaii p. 69-72. [doi: 10.1145/3123024.3123093]
20. Bedri A, Li R, Haynes M, Kosaraju RP, Grover I, Prioleau T, et al. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017;1(3):1-20 [FREE Full text] [doi: 10.1145/3130902] [Medline: 30135957]

21. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Kelbaek H, Munck O, Christensen NJ, Godtfredsen J. Central haemodynamic changes after a meal. Br Heart J 1989;61(6):506-509 [FREE Full text] [doi: 10.1136/hrt.61.6.506] [Medline: 2757863]
23. Covidence - Better systematic review management. Melbourne, Australia: Covidence URL: https://www.covidence.org/ [accessed 2023-07-09]
24. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
25. Hawley SK, Channer KS. Relative effects of fat-, carbohydrate- and protein-containing liquid diets on cardiac output in healthy adult subjects. Clin Sci (Lond) 1992;83(4):483-487. [doi: 10.1042/cs0830483] [Medline: 1330411]
26. Dagenais GR, Oriol A, McGregor M. Hemodynamic effects of carbohydrate and protein meals in man: rest and exercise. J Appl Physiol 1966;21(4):1157-1162 [FREE Full text] [doi: 10.1152/jappl.1966.21.4.1157] [Medline: 5916644]
27. Høost U, Kelbaek H, Rasmusen H, Court-Payen M, Christensen NJ, Pedersen-Bjergaard U, et al. Haemodynamic effects of eating: the role of meal composition. Clin Sci (Lond) 1996;90(4):269-276. [doi: 10.1042/cs0900269] [Medline: 8777833]
28. Waaler BA, Eriksen M. Post-prandial cardiovascular responses in man after ingestion of carbohydrate, protein or fat. Acta Physiol Scand 1992;146(3):321-327. [doi: 10.1111/j.1748-1716.1992.tb09426.x] [Medline: 1481689]
29. Shen Y, Arablouei R, de Hoog F, Xing H, Malan J, Sharp J, et al. In-situ fish heart-rate estimation and feeding event detection using an implantable biologger. IEEE Trans Mob Comput 2023;22(2):968-982. [doi: 10.1109/tmc.2021.3086496]
30. Shen Y, Arablouei R, Hoog FD, Malan J, Sharp J, Shouri S, et al. Estimating heart rate and detecting feeding events of fish using an implantable biologger. Piscataway, New Jersey: IEEE; 2020 Presented at: Proceedings of the 2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN); April 21-24, 2020; Sydney, NSW, Australia p. 37. [doi: 10.1109/ipsn48710.2020.00-48]
31. Hotta S, Mori T, Uchida D, Maeda K, Yaginuma Y, Inomata A. Eating moment recognition using heart rate responses. New York, NY, United States: ACM; 2017 Presented at: Proceedings of the 2017 ACM Int Jt Conf Pervasive Ubiquitous Comput Proc 2017 ACM Int Symp Wearable Comput; September 11-15, 2017; Maui, Hawaii p. 69-72. [doi: 10.1145/3123024.3123093]
32. Warwick PM. Thermic and glycemic responses to bread and pasta meals with and without prior low-intensity exercise. Int J Sport Nutr Exerc Metab 2007;17(1):1-13. [doi: 10.1123/ijsnem.17.1.1] [Medline: 17460329]
33. Calcagno M, Kahleova H, Alwarith J, Burgess NN, Flores RA, Busta ML, et al. The thermic effect of food: a review. J Am Coll Nutr 2019;38(6):547-551. [doi: 10.1080/07315724.2018.1552544] [Medline: 31021710]
34. Scott CB, Devore R. Diet-induced thermogenesis: variations among three isocaloric meal-replacement shakes. Nutrition 2005;21(7-8):874-877. [doi: 10.1016/j.nut.2004.12.010] [Medline: 15975497]
35. Mostardi R, Kubica R, Veicsteinas A, Margaria R. The effect of increased body temperature due to exercise on the heart rate and on the maximal aerobic power. Eur J Appl Physiol Occup Physiol 1974;33(3):237-245. [doi: 10.1007/BF00421151] [Medline: 4442403]
36. Taelman J, Vandeput S, Spaepen A, Van Huffel S. Influence of Mental Stress on Heart Rate Heart Rate Variability. Berlin, Heidelberg: Springer; 2009 Presented at: Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering; November 23-27, 2008; Antwerp, Belgium p. 1366-1369. [doi: 10.1007/978-3-540-89208-3_324]
37. Kappeler-Setz C, Gravenhorst F, Schumm J, Arnrich B, Tröster G. Towards long term monitoring of electrodermal activity in daily life. Pers Ubiquit Comput 2011;17(2):261-271 [FREE Full text] [doi: 10.1007/s00779-011-0463-4]
38. Caruelle D, Gustafsson A, Shams P, Lervik-Olsen L. The use of electrodermal activity (EDA) measurement to understand consumer emotions – a literature review and a call for action. Journal of Business Research 2019;104:146-160. [doi: 10.1016/j.jbusres.2019.06.041]
39. Chikwetu L, Miao Y, Woldetensae MK, Bell D, Goldenholz DM, Dunn J. Does deidentification of data from wearable devices give us a false sense of security? A systematic review. Lancet Digit Health 2023;5(4):e239-e247 [FREE Full text] [doi: 10.1016/S2589-7500(22)00234-5] [Medline: 36797124]


## Abbreviations

**EDA:** electrodermal activity
**IMU:** inertial measurement units
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

*Edited by T de Azevedo Cardoso; submitted 24.08.23; peer-reviewed by KY Hsieh, S Okita; comments to author 03.04.24; revised version received 23.04.24; accepted 24.07.24; published 10.09.24.*

*Please cite as:*
*Chikwetu L, Vakili P, Takais A, Younes R*
*Central Hemodynamic and Thermoregulatory Responses to Food Intake as Potential Biomarkers for Eating Detection: Systematic Review*
*Interact J Med Res 2024;13:e52167*
*URL: https://www.i-jmr.org/2024/1/e52167*
*doi:10.2196/52167*
*PMID:*

Review

# Improving Triage Accuracy in Prehospital Emergency Telemedicine: Scoping Review of Machine Learning–Enhanced Approaches

Daniel Raff[1], MD; Kurtis Stewart[2], MSc; Michelle Christie Yang[2], BSc; Jessie Shang[2]; Sonya Cressman[2,3], PhD; Roger Tam[4,5], PhD; Jessica Wong[6], MSc; Martin C Tammemägi[7], PhD; Kendall Ho[2], MD

[1]Department of Family Practice, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

[2]Department of Emergency Medicine, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

[3]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[4]School of Biomedical Engineering, Faculty of Applied Science, The University of British Columbia, Vancouver, BC, Canada

[5]Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

[6]Computer Science, Faculty of Science, The University of British Columbia, Vancouver, BC, Canada

[7]Faculty of Applied Health Sciences, Brock University, St. Catharines, ON, Canada

**Corresponding Author:**
Kendall Ho, MD
Department of Emergency Medicine
Faculty of Medicine
The University of British Columbia
818 W10 Ave
3rd floor
Vancouver, BC, V5Z1M9
Canada
Phone: 1 6048220327
Email: kendall.ho@ubc.ca

## Abstract

**Background:** Prehospital telemedicine triage systems combined with machine learning (ML) methods have the potential to improve triage accuracy and safely redirect low-acuity patients from attending the emergency department. However, research in prehospital settings is limited but needed; emergency department overcrowding and adverse patient outcomes are increasingly common.

**Objective:** In this scoping review, we sought to characterize the existing methods for ML-enhanced telemedicine emergency triage. In order to support future research, we aimed to delineate what data sources, predictors, labels, ML models, and performance metrics were used, and in which telemedicine triage systems these methods were applied.

**Methods:** A scoping review was conducted, querying multiple databases (MEDLINE, PubMed, Scopus, and IEEE Xplore) through February 24, 2023, to identify potential ML-enhanced methods, and for those eligible, relevant study characteristics were extracted, including prehospital triage setting, types of predictors, ground truth labeling method, ML models used, and performance metrics. Inclusion criteria were restricted to the triage of emergency telemedicine services using ML methods on an undifferentiated (disease nonspecific) population. Only primary research studies in English were considered. Furthermore, only those studies using data collected remotely (as opposed to derived from physical assessments) were included. In order to limit bias, we exclusively included articles identified through our predefined search criteria and had 3 researchers (DR, JS, and KS) independently screen the resulting studies. We conducted a narrative synthesis of findings to establish a knowledge base in this domain and identify potential gaps to be addressed in forthcoming ML-enhanced methods.

**Results:** A total of 165 unique records were screened for eligibility and 15 were included in the review. Most studies applied ML methods during emergency medical dispatch (7/15, 47%) or used chatbot applications (5/15, 33%). Patient demographics and health status variables were the most common predictors, with a notable absence of social variables. Frequently used ML models included support vector machines and tree-based methods. ML-enhanced models typically outperformed conventional triage algorithms, and we found a wide range of methods used to establish ground truth labels.

**Conclusions:** This scoping review observed heterogeneity in dataset size, predictors, clinical setting (triage process), and reported performance metrics. Standard structured predictors, including age, sex, and comorbidities, across articles suggest the importance of these inputs; however, there was a notable absence of other potentially useful data, including medications, social variables, and health system exposure. Ground truth labeling practices should be reported in a standard fashion as the true model performance hinges on these labels. This review calls for future work to form a standardized framework, thereby supporting consistent reporting and performance comparisons across ML-enhanced prehospital triage systems.

***(Interact J Med Res 2024;13:e56729)*** doi:10.2196/56729

**KEYWORDS**

telemedicine; machine learning; emergency medicine; artificial intelligence; chatbot; triage; scoping review; prehospital

## Introduction

Surging emergency department (ED) visits lead to overcrowding in the ED setting, contributing to adverse patient outcomes, staffing challenges, and health system constraints [1]. Challenges in maintaining ED capacity are estimated to cost millions in health care expenditures [2]. Various interventions have been proposed to improve ED conditions, and prehospital telemedicine triage systems are among the most promising, having the potential to prioritize patients based on their likelihood of requiring emergency versus community-based care [3], potentially alleviating the influx of low-acuity patients that would otherwise be managed in the high-cost, resource-intensive ED setting [4,5]. Such systems, including emergency medical dispatch, nurse-staffed telephone lines, and symptom checkers (chatbots), share the common goal to triage patients based on the information that is provided at the first contact for an urgent health concern. These prehospital services often do not include physician assessments, instead using either rule-based algorithms or health personnel for patient triage [6]. In telemedicine, defined as the delivery of health care services at a distance [7], the inherent scarcity of objective or physical measures such as vital signs has spurred efforts to improve risk prediction using machine learning (ML) models applied to a wide array of information sources such as free text from patient intake calls and vital symptoms monitoring [6,8].

Machine learning has been recognized as one option for improving the accuracy of prehospital telemedicine triage systems. To date, ML has commonly been applied in areas of precision medicine (ie, prediction of the success of treatment regimens), though it is rapidly expanding into diverse sectors of health care [9]. In the ED setting, ML models show promise in their ability to accurately predict inpatient admissions and sepsis [10,11]. Incorporating contextual information into ML models can improve prediction of prehospital emergency services [5]. However, there is a lack of understanding and evidence-based practices regarding how ML can optimally be implemented in remote prehospital settings as compared with more data-rich, in-person settings such as the ED [5,10,11].

In supervised learning, ML models learn from labeled data that serve as the "ground truth." Ground truth refers to the nature of the problem that is the target of the ML model; in the context of prehospital triage, ground truth is the "correct" triage outcome of a patient. The exact process for defining ground truth is complex and substantially varies across studies. Unlike clear binary ground truths, such as "alive" or "deceased," determining the "correct" triage outcome is more complex, potentially involving subsequent chart reviews or nuanced clinician assessments, thus introducing ambiguity.

A review of the predictors, labels, and models used in prehospital triage systems needs to establish an evidence base for efficient ML methods of triaging patients and identification of gaps not included in the existing models. Without a review, it is unclear whether existing methods and models may generalize to certain settings or be free from biases [12]. Thus, we conducted a scoping review to support our understanding of ML applications in remote prehospital settings. Scoping reviews are useful in the research fields of social sciences and health care, especially when a topic has not yet been comprehensively reviewed or exhibits a large, complex, or heterogeneous nature not amenable to a more thorough systematic review [13]. While there have been literature reviews outlining ML methods in emergency settings [14-16], no specific review touches on prehospital telemedicine triage services. Related reviews such as Sánchez-Salmerón et al [15] focused on in-person triage, as opposed to prehospital and remote triage. Through this scoping review, we aim to explore what evidence exists to compare the effectiveness of ML-enabled strategies with conventional triage methods in improving outcomes for patients seeking care through telemedicine services. We also aim to explore (1) what data sources and approaches are used for extracting meaningful predictors and labels, (2) what models and performance metrics are used, and (3) the processes of telemedicine triage systems where ML has been applied.

## Methods

### Search Strategy

Elements of the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) protocol [17] and the Population, Intervention, Comparison, and Outcome (PICO) framework [18,19] were used to guide our search strategy. The search included all articles published prior to February 24, 2023. Methodological frameworks for scoping reviews from Arksey and O'Malley [20] and Levac et al [21] were followed.

The search strategy began with initial searches conducted through MEDLINE to extract terms based on article titles and abstracts. Keywords and expressions included both regular and Medical Subject Headings (MeSH) terms. The following search expression was developed and applied: "Telephone/ or telephone.mp. OR phone.mp. OR telemedicine.mp. OR

Telemedicine/ OR Emergency Medical Service Communication Systems/ OR Emergency Medical Dispatch/ OR dispatch.mp. OR hotline.mp. OR Hotlines/ OR prehospital.mp. OR pre-hospital.mp. OR remote consultation.mp. OR Remote Consultation/" AND "machine learning.mp. OR Artificial Intelligence/ OR Machine Learning/ OR artificial intelligence.mp. OR natural language processing.mp. OR Natural Language Processing/ OR chatbot.mp." AND "triage/ OR triage.mp"

Our search followed the 3 phases listed from the PRISMA-ScR protocol: identification, screening, and inclusion [17] (Multimedia Appendix 1). Using the search expression, articles were retrieved from the following databases: MEDLINE, PubMed, Scopus, and IEEE Xplore. These databases were selected for this review due to their ability to cover the most scientific information in fields such as telemedicine and health care decision-making while also being previously used by several health and technology-related reviews [15,16,22]. In order to limit bias and ensure reproducibility, we elected to exclusively include articles identified through the predefined search strategy.

## Article Selection

All identified records were combined (n=296) and duplicates (n=131) removed resulting in 165 records remaining. The results of the initial search after duplicate removal are found in Multimedia Appendix 2. Articles needed to meet a set of inclusion and exclusion criteria (Table 1), which were developed to ensure that articles were relevant to prehospital telemedicine triage, conducted remotely as opposed to in the ED. During an initial piloting phase, 3 researchers (DR, KS, and MY) piloted these criteria to ensure consistency across them. Each researcher independently screened each title based on the inclusion and exclusion criteria, with discrepancies reaching consensus through discussion, which resulted in 32 records remaining (n=133 excluded). Articles that passed the title screening were then screened for relevant abstracts following a similar process resulting in 15 records remaining (n=17 excluded). The remaining articles were then read in their entirety and all 15 records were deemed eligible and included in the results of the review. Of the articles excluded, most were excluded for the provision of in-person care (as opposed to data collected remotely), were not primary research studies, or were not for general purpose triage (eg, stroke-specific triage).

**Table 1.** Inclusion and exclusion criteria of screening strategy.

| PICO[a]/other element | Inclusion criteria | Exclusion criteria |
|---|---|---|
| Population | Undifferentiated population seeking emergency services (including COVID-19 assessments). | Specialty-specific population (eg, stroke, heart disease). |
| Concept | Triage of emergency telemedicine services enhanced by any ML[b] method that includes only data collected remotely. | Only conventional triage methods used or ML models that include predictors derived from physical assessments (eg, vital signs). Internet of medical things devices requiring physical or in-person assessments were excluded (eg, home blood pressure equipment). |
| Context | Provision of emergency telemedicine services, including web-based symptom checkers, clinician-staffed telephone line, or emergency medical dispatch. | Provision of in-person emergency care. |
| Evidence type | Primary research studies. | Literature reviews, protocols, guidelines, letters, gray literature, and qualitative studies excluded. |
| Language | Studies published in English. | Studies in languages beyond English. |
| Date | Published before February 24, 2023. | Published on or after February 24, 2023. |

[a]PICO: Population, Intervention, Comparison, and Outcome.

[b]ML: machine learning.

## Data Extraction

We extracted data from the included articles with an aim to understand how research in this domain is conducted. A data extraction tool was developed using the JBI Manual for Evidence Synthesis template [23] with a pilot step on 2 sources conducted by 2 researchers (JS and DR). The data of interest fell in four main categories:

1. Study characteristics: Author, year of publication, country of origin, the aim of the study, the result of the study, population assessed, dataset source, dataset size (the number of patient records), prospective/retrospective/deployed, and triage process.
2. ML model predictors and labels: Number of predictors and data types; methods for determining ground truth labels.
3. ML techniques and corresponding performance metrics:
4. Comparators, dataset partitions, data-preprocessing methods, performance metrics and values, and data quality analysis approaches.
5. Resources for future ML model development: Source code availability and software packages used.

This list of extraction items was supported by prior literature that involved emergency care–related triage [22,24] and was refined based on discussions among the authors. For each of the eligible articles included in this review, 1 author (JS) extracted and tabulated the relevant information, and the information was validated by another author (DR) with discrepancies further assessed by a last author (KS). All extracted information was analyzed by the authors to derive a narrative synthesis of the findings. Aligned with the scoping

review methodology, articles were not assessed for quality or risk of bias and no statistical analyses were conducted. The data extraction tool and raw data extracted are available in Multimedia Appendix 2.

### Triage Process

We found that the studies spanned 3 prehospital triage processes: emergency dispatch, telephone lines, and chatbot applications. When distinguishing between emergency dispatch and telephone line systems, context of the call and nature of interaction were the 2 main points of consideration. Emergency dispatch calls (eg, 9-1-1) are received and handled by trained emergency staff focused on collecting critical information on the emergency, such as the nature of the incident, location, and immediate risks. The dispatcher then makes decisions based on this information to allocate appropriate resources, such as ambulances or first responders. Conversely, telephone lines such as nurse-led helplines or crisis hotlines are designed to offer support, advice, and clinical guidance to individuals seeking health care information or experiencing a crisis [25]. The interaction is often more conversational and supportive, resulting in a more complex triage process that depends on the nature of the call, the expertise of the health care professionals involved, and the available resources or referrals.

### Predictors

We classified the predictor variables into 4 domains (demographic factors, operational characteristics, clinical factors, and unstructured data such as free text) and extracted how these data were handled in the models. These domains were determined based on the prehospital triage field and iteratively refined as data were extracted.

### Ground Truth Labels

Variability existed in how ground truth labeling methods were coded for ML processing due to diverse sources of training data and annotation methods. In our classification of the observed ground truth methods, we first identified whether the label was derived from (1) the usual clinical process, that is, subsequent clinician triage, or (2) outside of the usual clinical process, that is, post hoc. Within category 1, we further identified whether (1A) remote-only data were used, or (1B) physical, in-person data were used. Within category 2, we further classified whether (2A) labels were automatically derived or (2B) labels were manually derived. Automatically (systematically) derived labels are the results of a uniform and scalable application of a label-deriving algorithm across a dataset.

## Results

### Characteristics of the Studies

Figure 1 shows a summary of the search protocol phases followed for this review using the PRISMA-ScR flow diagram. Of the 15 unique studies included, 33% (5/15) of the articles were published in 2022, 26% (4/15) in 2021, 20% (3/15) in 2020, and the remaining 20% (3/15) of the articles between 2014 and 2019. The most frequently occurring country of publication was the United States (4/15, 27%), followed by Japan (2/15, 13%), with the remaining being published in 9 distinct countries (Table 2).

Most studies (11/15, 73%) were retrospective, using historical patient outcomes to assess the performance of ML models in triage prediction. Among the retrospective articles, 2 reported a combination of retrospective results and the performance of deployed models. Of the other 4 nonretrospective articles, 13% (2/15) focused solely on deployed models and 13% (2/15) carried out a prospective study to explore the applications of ML triage in a specific health care setting.

Moreover, 47% (7/15) of articles investigated the use of ML in emergency dispatch calls, 33% (5/15) of articles focused on chatbot-style applications that could be accessed via the internet or within the ED, and 20% (3/15) of articles investigated ML in telephone lines, including 13% (2/15) nurse-led phone lines and 7% (1/15) crisis hotline.

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

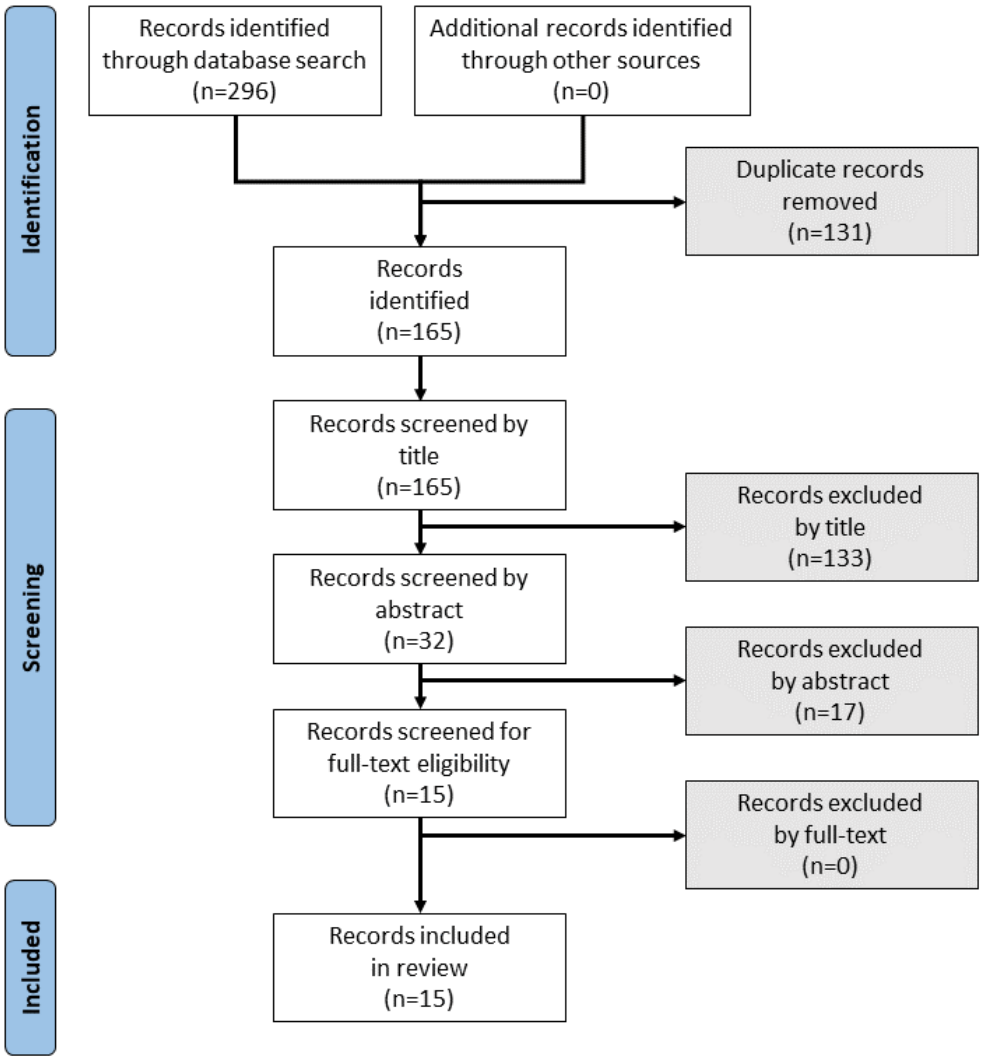


**Table 2.** List of 15 studies included in the review and their characteristics.

| Author (year) | Study setting (country) | Dataset size[a] | Labels included | Aim | Triage phase |
|---|---|---|---|---|---|
| Anthony et al (2021) [26] | South Africa | 93 | 3 binary | Classify critical conditions in emergency calls | Dispatch |
| Ceklic et al (2022) [27] | Australia | 11,971 | 1 binary | Classify severity of traffic crash incidents in dispatch calls | Dispatch |
| Chin et al (2022) [28] | Taiwan | 114 | 1 binary | Classify severity of traffic crash incidents in dispatch calls | Dispatch |
| Cotte et al (2022) [29] | Germany | 385 | 1 multiclass | Classify triage decisions using a symptom assessment app | Chatbot |
| Ferri et al (2021) [30] | Spain | 1,244,624 | 3 binary and multiclass | Classify emergency incidents in dispatch calls | Dispatch |
| Gatto et al (2022) [31] | United States | 574 | 1 binary | Classify severity in patient's text-based inquiries | Chatbot |
| Inokuchi et al (2022) [32] | Japan | 15,442 | 1 binary | Identify undertriage in prehospital telephone triage | Nurse-led phone line |
| Lai et al (2020) [33] | United States | —[b] | — | Classify triage for prehospital COVID-19 cases | Chatbot |
| Marchiori et al (2021) [25] | Switzerland | >900,000[c] | 1 multiclass | Evaluate AI[d]-powered chatbot for symptom-checker triage | Chatbot |
| Morse et al (2020) [34] | United States | 26,646 | — | Evaluate AI-powered chatbot for symptom-checker triage | Chatbot |
| Pacula et al (2014) [35] | United States | 427 | 2 multiclass | Classify triage and distress indicators in crisis hotline chats | Crisis hotline |
| Spangler et al (2019) [3] | Sweden | 68,668 | 1 continuous | Validate ML[e]-generated risk scores for prehospital care | Dispatch (operated by nurses) |
| Tollinton et al (2020) [5] | United Kingdom | 1,188,509 | 1 binary | Classify triage of unconscious patients in dispatch calls | Dispatch |
| Veladas et al (2021) [36] | Portugal | 269,669 | 1 multiclass | Classify clinical pathways from text data | Nurse-led phone line |
| Yunoki et al (2014) [4] | Japan | 61,927 | 1 multiclass | Classify triage categories from phone call data | Dispatch |

[a]Number of patient records included.

[b]Dataset size used for model development was not stated for this study or information on labels was not included.

[c]The study stated that "more than 900,000 case records" were included.

[d]AI: artificial intelligence.

[e]ML: machine learning.

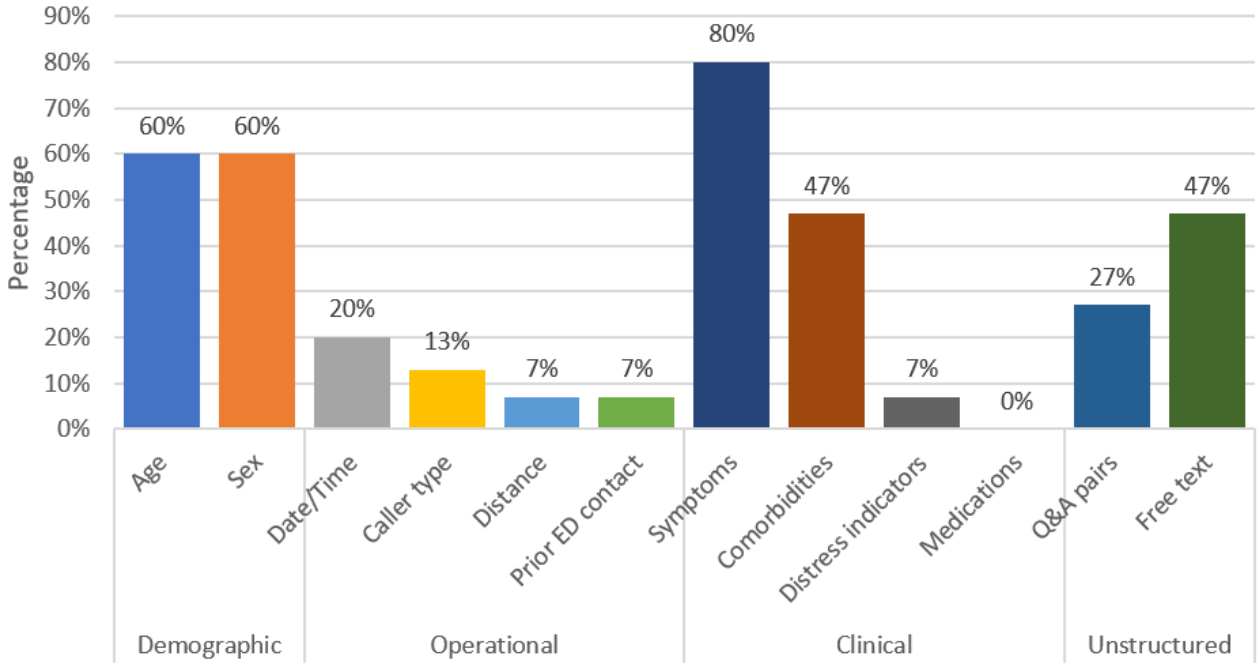
The size of the dataset used across studies varied considerably, with a median sample size of 21,044 observations. The largest dataset comprised 1,224,624 anonymized patient records [30] and the smallest dataset included 93 call transcripts [28]. Distinct methods were reported to handle missing data; for example, Ferri et al [30] excluded all call records with missing values, while Inokuchi et al [32] performed imputation to account for missing data (3918/19,114, 20.5% of cases had missing data) using the k-nearest neighbors (k-NN) algorithm.

## Predictors

Several articles used predictors that spanned multiple categories, as exemplified by Ferri et al [30], who extracted patient demographic data (age, gender), operational data (date, caller type), and unstructured data (clinical free-text observations). In contrast, Chin et al [28] used unstructured data (dispatch call transcript) as the only data source in their ML model development, citing the "higher expressiveness of patient condition than structured data." The frequency of occurrence for each of these domains is shown in Figure 2.

The mapping and processing techniques varied across data types. Most commonly, many input variables were one-hot encoded, such as age groupings, sex, day of week, clinical indicators (symptoms, comorbidities), and question-answer pairs. Notably, distance to ED was one of the few continuous predictors used [3]. For unstructured data, various natural language processing (NLP) techniques were employed to transform these data to useful inputs, such as bag-of-words methods or text vectorization, which are strategies to transform text data into a numerical representation that can be processed by ML models (see Multimedia Appendix 3 for more information).

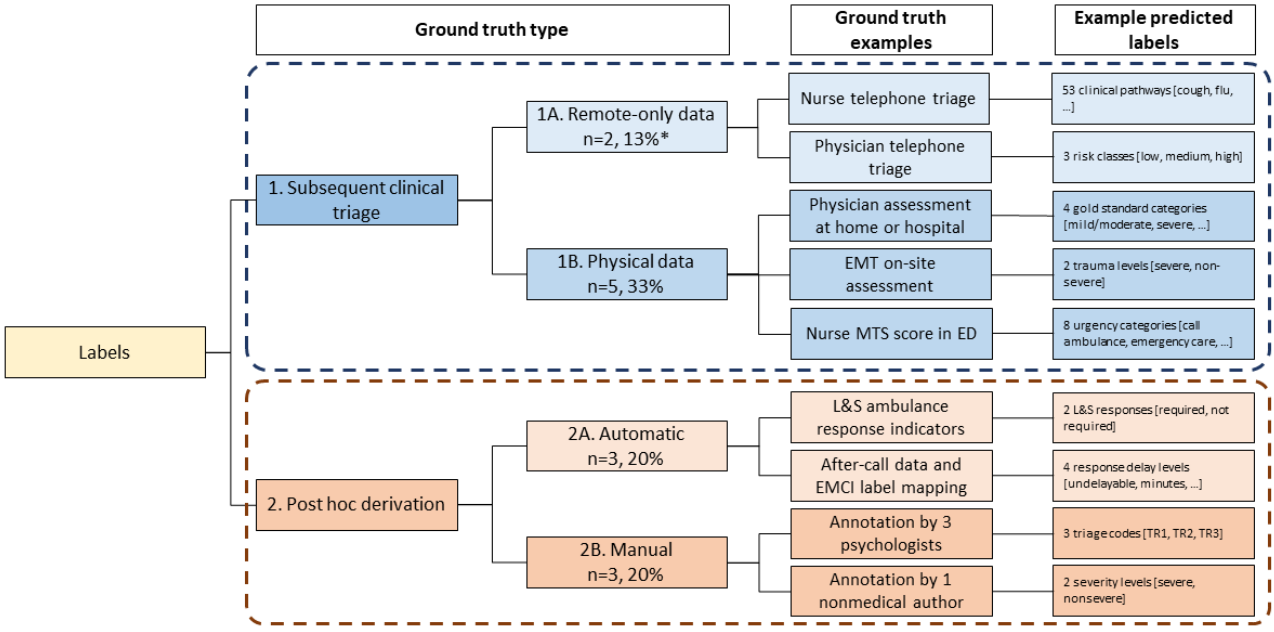
**Figure 2.** Percentage of predictor types included in the studies. ED: emergency department; Q&A: question and answer.




## Data Labeling: Determining Ground Truth

Two studies (2/15, 13%) [33,34] did not provide sufficient information to determine their ground truth labeling method. Most of the ground truth methods (5/15, 33%) used physical assessments from the usual clinical process; however, within this group input data were from 3 different types of providers (nurses, physicians, and ambulance crew) with each employing a different triage or categorization system (Figure 3) [25]. This results in variability as similar information is outputted with different labels, which limits comparability (eg, patient conveyance [5] or "gold standard" ED triage protocols [4]). Notably, the 3 (20%) studies [3,27,30] that used automatic post hoc (2A) methods had larger sample sizes (mean 4,32,792 observations; range 15,550-1,244,624) versus studies [26,31,35] using manual (2B) methods (mean 364 observations; range 93-573). Across the 4 methods, 6 studies (40%) predicted labels that were binary and based on symptom or condition severity (eg, life-threatening [30], severe trauma [28]), while 7 studies (47%) predicted multiclass labels (eg, 1 of 53 clinical pathways [36]).

**Figure 3.** Label and ground truth methodologies by annotation method. *Marchiori et al used remote physician triage as ground truth for model training but manually derived labels for model testing. To ensure consistency in article classification, we have categorized all articles based on the ground truth method for model training. ED: emergency department; EMCI: emergency medical call incident; EMT: emergency medical technician; L&S: lights and sirens; MTS: Manchester Triage System.



### Interrater Reliability

Methods to enhance interrater reliability were rarely employed, in contrast to published guidelines concerning manual data labeling [37]. Marchiori et al [25] did so by ensuring the adequate training of data collectors by selecting "only records generated by top-ranked doctors (based on years of experience and internal audits)." In the study by Gatto et al [31], data annotators "invested significant time to educate themselves on the symptoms" observed in the dataset; however, it was noted "that the use of nonmedical professionals limited the degree of granularity with which the dataset was labelled." Pacula et al [35] used consensus among 3 evaluators when labeling data and measured the interannotator agreement using the Fleiss kappa statistic to confirm good agreement.

### ML Algorithms

Most studies applied more than 1 ML technique to method development (11/15, 73%). However, 2 studies (13%) did not disclose the ML algorithm used in their triage prediction model. Three (20%) studies published the source code for various stages of their ML model development and validation (Table 3).

Overall, the most popular ML algorithm was support vector machine (SVM; 8/15, 53%) and tree-based methods, including random forest (RF; 6/15, 40%) and extreme gradient boosting (XGBoost; 4/15, 27%). Other popular algorithms were neural networks (NN; 5/15, 33%), with various implementations observed, such as deep NNs, bidirectional long short-term memory models, and ensemble of deep learning networks (ie, the model employed by Ferri et al [30] is described as a "Deep Ensemble Multitask Classifier for Emergency Medical Calls" composed of 4 different subnetworks). Naïve Bayes (NB) and k-NN were similarly popular, followed by regression methods.

**Table 3.** Type and frequency of ML algorithms used by the included studies.

| ML[a] algorithm | Studies, n | Authors (citations) |
|---|---|---|
| Support vector machine | 8 | Anthony et al [26], Ceklic et al [27], Chin et al [28], Gatto et al [31], Inokuchi et al [32], Pacula et al [35], Spangler et al [3][b], and Veladas et al [36] |
| Random forest | 6 | Anthony et al [26], Ferri et al [30], Inokuchi et al [32], Tollinton et al [5], Spangler et al [3], and Veladas et al [36] |
| Neural networks | 5 | Ceklic et al [27], Ferri et al [30], Gatto et al [31], Inokuchi et al [32], and Marchiori et al [25] |
| Extreme gradient boosting (XGBoost) | 4 | Ferri et al [30], Inokuchi et al [32], and Spangler et al [3] |
| Naïve Bayes | 4 | Ceklic et al [27], Chin et al [28], Ferri et al [30], and Veladas et al [36] |
| K-nearest neighbors | 4 | Anthony et al [26], Ceklic et al [27], Chin et al [28], and Gatto et al [31] |
| Logistic regression | 3 | Anthony et al [26], Ferri et al [30], and Spangler et al [3] |
| Bayesian network | 2 | Cotte et al [29] and Yunoki et al [4] |
| Decision tree[c] | 2 | Chin et al [28] and Tollinton et al [5] |
| Ensemble[d] | 2 | Ceklic et al [27] and Ferri et al [30] |
| LASSO[e] regression | 1 | Inokuchi et al [32] |
| Multilayer perceptron | 1 | Chin et al [28] |
| Hidden Markov model | 1 | Pacula et al [35] |
| Hierarchical attention network | 1 | Gatto et al [31] |
| Transformer-based | 1 | Gatto et al [31] |
| Unspecified algorithms | 2 | Lai et al [33] and Morse et al [34] |

[a]ML: machine learning.

[b]Spangler et al [3] did not include support vector machine, random forest, neural networks, or regression models in their "Methods" but stated in their "Discussion" that they investigated these algorithms.

[c]Decision trees here include methods such as gradient boosting but not random forest or extreme gradient boosting.

[d]Ceklic et al [27] did not provide information about the specific models incorporated into their ensemble. Ferri et al [30] built their ensemble using a collection of deep learning subnetworks.

[e]LASSO: least absolute shrinkage and selection operator.

### Model Performance Metrics

$F_1$-score was the most used metric (6/15, 40%) to measure model performance, closely followed by accuracy (5/15, 33%) and area under the curve (AUC; 4/15, 27%). Furthermore, many studies tangentially used other quantitative measures, such as sensitivity or recall and specificity or precision. In the case where AUC scores were similar between the RF and XGBoost models in the study by Tollinton et al [5], analysis of other metrics showed that XGBoost outperformed RF models in terms of sensitivity (0.93 vs 0.62 in the combined model) but had lower specificity (0.17 vs 0.56 in the combined model).

Moreover, several studies investigated how different combinations of predictors and NLP techniques affected model performance across ML techniques. For instance, Pacula et al [35] present an interesting discussion on the impact of various approaches to dialogue processing (turn-level classification and accounting for speaker role) on triage prediction outcomes.

A summary of the comparators and top-performing models for each study, alongside information on evaluation metrics used and data train or test split, is shown in Table 4. For models that were evaluated against non–ML-enhanced methods of triage, those established triage systems are indicated as well.

**Table 4.** Top-performing ML algorithms used in each study and corresponding performance metric and model training or testing split grouped by triage process.

| Triage process and study | Top-performing model | Model comparators | Triage comparators | Performance metrics | Training data split (%) |
|---|---|---|---|---|---|
| **Emergency medical dispatch** | | | | | |
| Anthony et al [26] | SVM[a] | LR[b], RF[c], k-NN[d] | —[e] | Accuracy | 80 |
| Ceklic et al [27] | Ensemble | k-NN, SVM, NB[f], deep NN +NLP[g] | SJ-WA[h] dispatch | PPV[i], sensitivity (recall), $F_1$-score | 60 |
| Chin et al [28] | Bernoulli NB | k-NN, DT[j], SVM, NB +NLP | Dispatcher evaluation | Accuracy, PPV, NPV[k], sensitivity (recall), specificity | 91 |
| Ferri et al [30] | Ensemble | NB, LR, RF, GB[l] +NLP | Clinical decision tree | Accuracy, sensitivity (recall), PPV, $F_1$-score | 80 |
| Spangler et al [3] | XGBoost[m] | SVM, LR, RF, deep NN | Dispatch priority, National Early Warning Scores | AUC[n], PPV, sensitivity (recall) | 66 |
| Tollinton et al [5] | GB/RF[o] | — | — | AUC, sensitivity (recall), specificity | 80 |
| Yunoki et al [4] | BN[p,q] | — | — | Accuracy | 90 |
| **Symptom checker/chatbot[r]** | | | | | |
| Cotte et al [29] | BN[q] | — | MTS[s] | Cohen κ | N/A[t] |
| Gatto et al [31] | SBERT[u] | BERT[v], SVM, HAN[w], bi-LSTM[x] +NLP | — | PPV, sensitivity (recall), $F_1$-score | 80 |
| Marchiori et al [25] | Bi-LSTM | Convolutional NN, recurrent NN +NLP | — | PPV, sensitivity (recall), $F_1$-score | 60 |
| **Telephone line** | | | | | |
| Inokuchi et al [32] | RF[q] | LASSO[y] regression, deep NN, XGBoost | — | Area under the receiver operating characteristic curve, PPV, NPV, sensitivity (recall), specificity | 70 |
| Pacula et al [35] | SVM[q] | HMM[z] +NLP | — | AUC, $F_1$-score | 82 |
| Veladas et al [36] | SVM | RF, NB +NLP | — | Accuracy, $F_1$-score, PPV, sensitivity (recall) | 64 |

[a]SVM: support vector machine.

[b]LR: logistic regression.

[c]RF: random forest.

[d]K-NN: K-nearest neighbors.

[e]No comparator.

[f]NB: naïve Bayes.

[g]"+NLP" indicates that model performance was evaluated across various natural language processing techniques.

[h]SJ-WA: St John Ambulance in Western Australia.

[i]PPV: positive predictive value (precision).

[j]DT: decision tree.

[k]NPV: negative predictive value.

[l]GB: gradient boosting.

[m]XGBoost: extreme gradient boosting.

[n]AUC: area under the curve.

[o]The gradient boosting model scored better on sensitivity, but specificity was lower than RF.

[p]BN: Bayesian network.

[q]Alternate ML models were not compared.

[r]Lai et al [33] and Morse et al [34] reported on symptom checker triage systems but did not provide details on the underlying ML model and comparators or development.

[s]MTS: Manchester Triage System.

[t]N/A: not applicable.

[u]SBERT: sentence bidirectional encoder representation from transformers.

[v]BERT: Bidirectional Encoder Representation From Transformers.

[w]HAN: hierarchical attention network.

[x]Bi-LSTM: bidirectional long short-term memory.

[y]LASSO: least absolute shrinkage and selection operator.

[z]HMM: hidden Markov model.

### Model Data Splitting

Thirteen studies (87%) reported some method for validation, including randomly partitioning data into training or testing sets (10/15, 67%) or training or validation or testing sets (3/15, 20%). Most studies used some form of cross-validation in the training sets for model construction (8/15, 53%). In addition, various types of resampling procedures were used, such as *k*-fold cross-validation, stratified *k*-fold cross-validation, repeated *k*-fold cross-validation, or repeated random test-train splits.

## Discussion

### Principal Findings

We reviewed 15 primary research studies of ML-enhanced triage models in prehospital telemedicine settings where patients potentially require emergency care. We found that ML-enhanced triage systems typically outperformed conventional triage ones; however, there is likely a bias toward publishing these types of positive findings [38]. While the reviewed studies exhibited several commonalities in terms of predictor types, ML algorithms tested, and performance metrics, there were key discrepancies in how data were sourced and processed, particularly with regard to annotating ground truth labels. We discuss these similarities and differences in context and how the new evidence presented here relates to ED overcrowding and prehospital triage.

### Predictor Variables

A critical limitation of prehospital telemedicine triage systems is the lack of access to objective measures of the patient condition typically obtained through physical assessment, vital signs being fundamental metrics of illness severity. As a result, our inclusion criteria ensured that no study had access to a physical assessment of their participants. Therefore, by investigating the range of predictors employed by the included articles, we identified how subjective and indirect indicators of patient condition are used in remote triage. Our review indicates that patient symptoms, age, sex, and comorbidities were the most frequently occurring predictor variables among structured data, with NLP techniques used to extract features from unstructured data. The structured and unstructured data were not combined in any of the included articles. This finding emphasizes the critical importance of considering these 4 clinical and demographic factors for prehospital telemedicine triage where physical assessment data are unavailable. The consistent use of patient symptoms, demographics, and comorbidities as predictor variables across all 3 triage processes reinforces that they are reliable indicators of a suitable triage outcome. However, while demographic factors were frequently used across the studies, only 2 specific factors, namely, age and sex, were considered.

Notably, we also highlight the absence of previously identified important inputs to ML models [12]. Race or ethnicity, region or geography, medication history, and health system exposure (hospitalizations, etc) were not represented in the corpus; however, they should be taken into consideration in prehospital triage both to improve performance and ensure algorithmic fairness [39]. The absence of these variables both as predictors and for post hoc evaluation of algorithmic bias suggests a significant gap in the extant literature and motivates further investigation into their potential to generate accurate and equitable triage outcomes for more diverse populations. Continuing to develop a comprehensive list of the most significant variables driving remote emergency triage is invaluable to improving equitable patient outcomes for all populations [12]. Identifying common patterns in predictor selection (and what is absent) can inform the development of standardized guidelines for building ML algorithms for triage using remote-only data. A future with consistent measurements of physiologic metrics, such as vital signs, would also be invaluable to strengthen prediction. Developing systems or technology whereby these critical data can be captured remotely is a future direction worth exploring.

### Ground Truth Labeling Methods

Our review critically spotlights the variety of methods for annotating data labels from ground truth. Two papers did not provide sufficient information, and we classified the remaining 13 into 1 of 4 distinct methods, reflecting varying ML development philosophies. The most common method—(2A) post hoc automatic derivation of data labels—indicates preference for collaboration among domain experts to reduce human subjectivity and implement large-scale data mapping. Notably, Ferri et al [30] used a panel of 17 physicians to develop a mapping system, which was then automatically applied to more than 1 million records. In contrast, using the (2B) method, Pacula et al [35] reported that 3 psychologists manually annotated each of the 427 records. While this review cannot comment on the comparative efficiencies of these specific methods, we note that assembling large datasets with ground truth labels is an arduous and expensive task; thus, there may be scalability benefits to automatic methods [40].

While we observed trade-offs in data labeling, establishing methodologically robust ground truth is of paramount

importance. Inadequate representation of ground truth can lead to misclassification issues within the ML model, thus reducing predictive accuracy on external data. This could result in unintended and potentially serious consequences, especially in the context of emergency care. The ground truth triage outcome has proved challenging to pinpoint, and this scoping review reveals 4 methods in determining it. However, each method exhibits limitations and potential manifestations of human subjectivity.

The accuracy of remote triage is inherently limited by the absence of physical data, even when carried out by individuals with domain expertise (eg, physicians). Furthermore, human bias and lack of experience can lead to either overtriage (ie, false positives; eg, sending low-acuity patients to the ED) or undertriage (ie, false negatives; eg, advising high-acuity patients to stay home), even in cases when physical data are available. To address this, the strategy of selecting data from only highly reputable clinicians [25] serves as an inspiration for developing a standardized evaluation system to identify qualified data annotators for health care settings. An additional method to enhance triage accuracy is inclusion of downstream triage with access to more comprehensive data, such as vital signs, detailed physical assessment, or a professional with higher-level training (eg, a physician conducting a home visit) [32].

The methodical classification of patients based on a predetermined list of outcomes minimizes the risk for interrater disagreement. However, relying solely on these selected outcomes, such as hospital admission or 2-day mortality [3], assumes that they are the only factors and of equal importance in determining the overall risks associated with the patient. This approach overlooks insights that can be revealed from a circumstantial and personalized analysis of patient condition. In addition, many of the inclusion studies selected highly specific outcomes, such as sepsis, myocardial infarction, and cardiac arrest [26], which are not generalizable to different remote triage processes. One potential solution, which no study in this review used, is using a weighted kappa index to consider different categories and disagreements and capture the rank magnitude of disagreement [41].

Retrospective evaluation of patient data by multiple annotators holds potential for highly accurate ground truth labels, as it is the most comprehensive approach observed in this review. However, we note that the included articles lack information regarding how evaluators determined such classifications. Subjectivity in human decision-making persists and variation in annotators' levels of training and resolution strategies employed leaves room for further research. Again, implementing a standardized evaluation system to determine a qualified pool of data annotators becomes crucial to ensure reliability in the annotation of ground truth triage labels for ML-enhanced remote triage [37].

This analysis of subjectivity in ML systems underscores the need for nuance regarding objectivity and ground truth. To enhance remote triage algorithm generalizability, comprehensive datasets must be used. This entails capturing data from each of the predictor variables outlined in this review, as well as race or ethnicity, region or geography, and medical history. Furthermore, standardized protocols for data labeling across different clinical settings would facilitate consistency and enable meaningful comparisons and analyses. There is a need for more well-defined and externally valid procedures for determining the ground truth in the context of highly intricate and unstructured data such as remote patient triage.

## ML Techniques

This review presents needed insights on the cutting-edge application of ML methods for remote prehospital triage systems: SVM, tree-based, and NN methods were commonly used. These observations are consistent with other reviews of ED-based implementations [16,42]. To our knowledge, this is the first review to provide a mapping between ML model development and 3 processes of triage in emergency medicine: emergency medical dispatch, chatbot or symptom checker, and telephone line. We uniquely charted the best-performing model of each article to 1 of the 3 triage processes and generated insight on the specific elements underpinning each model. By doing so, we provide insight into how predictive ML models can be developed for different stages of triage before physical assessment.

## Limitations

We did not directly compare the varied performance metrics such as the $F_1$-score, precision, recall, and AUC, as this could lead to misleading conclusions. This was not an objective of our review due to inconsistencies in reported metrics and variability in study designs. For instance, certain metrics can be optimized at the expense of other metrics; for example, precision and recall are inversely related. Not all metrics are universally applicable; the $F_1$-score is unsuitable for multiclass problems found in a proportion of included studies. In addition, $F_1$-score assumes false negatives and false positives are equally costly, an incorrect assumption in triage problems where undertriage due to a false negative would have serious consequences. While AUC is perhaps the optimal metric choice, it still has sensitivity to class imbalances, which varied across the included studies. Of the only 4 articles that reported AUC, the performance ranged from 0.64 for prediction of conveyance [5] to 0.88 [35]. A direct comparison among articles reporting AUC was avoided due to differences in study contexts, which included diverse triage labels, significant variations in sample sizes and populations, limited reporting on class imbalance, and the use of different validation strategies. As this review did not facilitate any quantitative analysis of ML model performance, insights into the accuracy of ML-enhanced triage compared with conventional remote triage, as well as the quantitative impact that ML triage may have on patient outcomes and overall health care systems, could not be derived and future systematic work is warranted.

## Implications for Future Work

ML-enhanced triage presents an opportunity to alleviate the burden on EDs and support patients' decision-making when seeking emergency versus community-based care. Based on the evidence synthesized here, our calls to the field are to determine a prioritized list of high-value predictor variables to consider, standardize ground truth labeling, and form a consensus on

validation methods used, such that different health systems can continuously learn from new developments. This review provides a foundation for developing guidelines, which will also create opportunities for comparison across studies to quantitatively assess the effectiveness and benefits of ML-enhanced triage regarding patient outcomes and health system performance. While reducing resources spent on overtriage is also a priority, it is equally important to focus on the likelihood of undertriage, as it poses a significant risk to patients. Therefore, precision, recall, and specificity rates, as well as algorithmic bias [39], must be carefully monitored and improved in model development to ensure that safety and effectiveness are balanced in ML-enhanced triage systems.

Accurate triage recommendations do not guarantee that patients will follow them [43-45]. In the discussion of using telephone lines or chatbots during the triage process, further investigation is warranted into how delivery methods affect patients' likeliness to adhere to the advice generated by ML-enhanced triage systems. Similarly, in the dispatch and telephone line triage processes, a question that arises is whether ML-generated triage results influence the decision-making process of dispatchers or providers. This area of research could provide insight on the most effective stage at which ML-based assistance should be introduced in the triage process. Exploring the interplay between ML-enhanced triage advice, patient behavior, and clinician decision-making will contribute to the optimization of prehospital telemedicine triage in emergency care.

## Conclusions

Our scoping review of 15 recent studies of ML-enhanced prehospital telemedicine triage systems observed heterogeneity in dataset size, predictors, clinical setting (triage process), and reported performance metrics. Consequently, a comparison of ML performance across articles was not feasible, and we note that identifying the most efficient and accurate ML-enhanced triage system is valuable for future development and model deployment in prehospital settings, where a standardized performance metric such as the AUC would be important to facilitate comparisons. Standard structured predictors, including symptoms, age, sex, and comorbidities, across articles suggest the importance of these inputs; however, there was a notable absence of other potentially useful data, including medications and health system exposure. With advancing technology of transformer-based models [46,47], there exists the potential for combining structured and unstructured data; an approach that was also absent in the included articles. The lack of social variables leaves the potential for algorithmic bias critically unexplored. Ground truth labeling practices should be reported in a standard fashion as the true model performance hinges on these labels. This review establishes an evidence base for future investigations and an opportunity to form a consensus and standardized framework, thereby supporting consistent reporting, performance comparisons, and collaboratively developed ML-enhanced prehospital triage systems.

## Authors' Contributions

DR, KS, SC, RT, JW, MCT, and KH developed the idea for the scoping review. DR and KS developed the search methodology. MCY conducted the literature search. MCY and JS wrote the initial draft. DR and KS completed the initial draft. All authors reviewed, edited, and approved the final manuscript.

## Conflicts of Interest

KH is the lead of the UBC Digital Emergency Medicine Unit, which was provided grant funding from the BC Ministry of Health to implement the HealthlinkBC Virtual Physicians program, the BC prehospital emergency telemedicine program.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 106 KB - ijmr_v13i1e56729_app1.docx ]

Multimedia Appendix 2
Data extraction tool and search results.
[ZIP File (Zip Archive), 196 KB - ijmr_v13i1e56729_app2.zip ]

Multimedia Appendix 3
Additional findings.
[DOCX File , 15 KB - ijmr_v13i1e56729_app3.docx ]

## References

1. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS One 2018;13(8):e0203316 [FREE Full text] [doi: 10.1371/journal.pone.0203316] [Medline: 30161242]
2. Gidora H, Borycki EM, Kushniruk AW. Effects of telenursing triage and advice on healthcare costs and resource use. Stud Health Technol Inform 2019;257:133-139. [Medline: 30741185]

3. Spangler D, Hermansson T, Smekal D, Blomberg H. A validation of machine learning-based risk scores in the prehospital setting. PLoS One 2019;14(12):e0226518 [FREE Full text] [doi: 10.1371/journal.pone.0226518] [Medline: 31834920]
4. Yunoki S, Hamagami T, Oshige K, Kawakami C, Suzuki N. High accuracy of call triage decision by Bayesian network. Electron Comm Jpn 2014;97(1):62-69. [doi: 10.1002/ecj.11439]
5. Tollinton L, Metcalf AM, Velupillai S. Enhancing predictions of patient conveyance using emergency call handler free text notes for unconscious and fainting incidents reported to the London ambulance service. Int J Med Inform 2020;141:104179 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104179] [Medline: 32663739]
6. Lidal IB, Holte HH, Vist GE. Triage systems for pre-hospital emergency medical services—a systematic review. Scand J Trauma Resusc Emerg Med 2013;21:28 [FREE Full text] [doi: 10.1186/1757-7241-21-28] [Medline: 23587133]
7. World Health Organization. Consolidated telemedicine implementation guide. 2022. URL: https://www.who.int/publications/i/item/9789240059184 [accessed 2023-07-24]
8. Bohm K, Kurland L. The accuracy of medical dispatch—a systematic review. Scand J Trauma Resusc Emerg Med 2018;26(1):94 [FREE Full text] [doi: 10.1186/s13049-018-0528-8] [Medline: 30413213]
9. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
10. King Z, Farrington J, Utley M, Kung E, Elkhodair S, Harris S, et al. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. NPJ Digit Med 2022 Jul 26;5(1):104. [doi: 10.1038/s41746-022-00649-y] [Medline: 35882903]
11. Adams R, Henry KE, Sridharan A, Soleimani H, Zhan A, Rawat N, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nat Med 2022;28(7):1455-1460. [doi: 10.1038/s41591-022-01894-0] [Medline: 35864252]
12. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
13. Bichu YM, Hansa I, Bichu AY, Premjani P, Flores-Mir C, Vaid NR. Applications of artificial intelligence and machine learning in orthodontics: a scoping review. Prog Orthod 2021;22(1):18 [FREE Full text] [doi: 10.1186/s40510-021-00361-9] [Medline: 34219198]
14. Shafaf N, Malek H. Applications of machine learning approaches in emergency medicine; a review article. Arch Acad Emerg Med 2019;7(1):34 [FREE Full text] [Medline: 31555764]
15. Sánchez-Salmerón R, Gómez-Urquiza JL, Albendín-García L, Correa-Rodríguez M, Martos-Cabrera MB, Velando-Soriano A, et al. Machine learning methods applied to triage in emergency services: a systematic review. Int Emerg Nurs 2022;60:101109. [doi: 10.1016/j.ienj.2021.101109] [Medline: 34952482]
16. Gao F, Boukebous B, Pozzar M, Alaoui E, Sano B, Bayat-Makoei S. Predictive models for emergency department triage using machine learning: a systematic review. Obstet Gynecol Res 2022;5(2):136-157. [doi: 10.26502/ogr085]
17. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Sackett DL. Evidence-based medicine. Semin Perinatol 1997;21(1):3-5. [doi: 10.1016/s0146-0005(97)80013-4] [Medline: 9190027]
19. Aslam S, Emmanuel P. Formulating a researchable question: a critical step for facilitating good clinical research. Indian J Sex Transm Dis AIDS 2010;31(1):47-50 [FREE Full text] [doi: 10.4103/0253-7184.69003] [Medline: 21808439]
20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
21. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
22. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JMC. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 2020;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
23. Micah DP, Christina G, Patricia M, Zachary M, Andrea CT, Hanan K. Chapter 11: scoping reviews. In: JBI Manual for Evidence Synthesis JBI. New York, NY: JBI; 2020.
24. Salman OH, Taha Z, Alsabah MQ, Hussein YS, Mohammed AS, Aal-Nouman M. A review on utilizing machine learning technology in the fields of electronic emergency triage and patient priority systems in telemedicine: coherent taxonomy, motivations, open research challenges and recommendations for intelligent future work. Comput Methods Programs Biomed 2021;209:106357. [doi: 10.1016/j.cmpb.2021.106357] [Medline: 34438223]
25. Marchiori C, Dykeman D, Girardi I, Ivankay A, Thandiackal K, Zusag M, et al. Artificial intelligence decision support for medical triage. AMIA Annu Symp Proc 2020;2020:793-802 [FREE Full text] [Medline: 33936454]
26. Anthony T, Mishra AK, Stassen W, Son J. The feasibility of using machine learning to classify calls to South African emergency dispatch centres according to prehospital diagnosis, by utilising caller descriptions of the incident. Healthcare (Basel) 2021;9(9):1107 [FREE Full text] [doi: 10.3390/healthcare9091107] [Medline: 34574881]

27. Ceklic E, Ball S, Finn J, Brown E, Brink D, Bailey P, et al. Ambulance dispatch prioritisation for traffic crashes using machine learning: a natural language approach. Int J Med Inform 2022;168:104886. [doi: 10.1016/j.ijmedinf.2022.104886] [Medline: 36306652]
28. Chin KC, Cheng YC, Sun JT, Ou CY, Hu CH, Tsai MC, et al. Machine learning-based text analysis to predict severely injured patients in emergency medical dispatch: model development and validation. J Med Internet Res 2022;24(6):e30210 [FREE Full text] [doi: 10.2196/30210] [Medline: 35687393]
29. Cotte F, Mueller T, Gilbert S, Blümke B, Multmeier J, Hirsch MC, et al. Safety of triage self-assessment using a symptom assessment app for walk-in patients in the emergency care setting: observational prospective cross-sectional study. JMIR Mhealth Uhealth 2022;10(3):e32340 [FREE Full text] [doi: 10.2196/32340] [Medline: 35343909]
30. Ferri P, Sáez C, Félix-De Castro A, Juan-Albarracín J, Blanes-Selva V, Sánchez-Cuesta P, et al. Deep ensemble multitask classification of emergency medical call incidents combining multimodal data improves emergency medical dispatch. Artif Intell Med 2021;117:102088 [FREE Full text] [doi: 10.1016/j.artmed.2021.102088] [Medline: 34127234]
31. Gatto J, Seegmiller P, Johnston G, Preum SM. Identifying the perceived severity of patient-generated telemedical queries regarding COVID: developing and evaluating a transfer learning-based solution. JMIR Med Inform 2022;10(9):e37770 [FREE Full text] [doi: 10.2196/37770] [Medline: 35981230]
32. Inokuchi R, Iwagami M, Sun Y, Sakamoto A, Tamiya N. Machine learning models predicting undertriage in telephone triage. Ann Med 2022;54(1):2990-2997 [FREE Full text] [doi: 10.1080/07853890.2022.2136402] [Medline: 36286496]
33. Lai L, Wittbold KA, Dadabhoy FZ, Sato R, Landman AB, Schwamm LH, et al. Digital triage: novel strategies for population health management in response to the COVID-19 pandemic. Healthc (Amst) 2020;8(4):100493 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100493] [Medline: 33129176]
34. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]
35. Pacula M, Meltzer T, Crystal M, Srivastava A, Marx B. Automatic detection of psychological distress indicators and severity assessment in crisis hotline conversations. : IEEE; 2014 Presented at: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2014 July 14; Florence, Italy p. 4863-4867. [doi: 10.1109/icassp.2014.6854526]
36. Veladas R, Yang H, Quaresma P, Gonçalves T, Vieira R, Sousa PC, et al. Aiding clinical triage with text classification. In: Marreiros G, Melo FS, Lau N, Lopes Cardoso H, Reis LP, editors. Progress in Artificial Intelligence. Cham: Springer International Publishing; 2021:83-96.
37. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med 2022;5(1):2 [FREE Full text] [doi: 10.1038/s41746-021-00549-7] [Medline: 35013569]
38. Ioannidis JPA. Why most published research findings are false. PLoS Med 2005;2(8):e124 [FREE Full text] [doi: 10.1371/journal.pmed.0020124] [Medline: 16060722]
39. Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng 2023;7(6):719-742 [FREE Full text] [doi: 10.1038/s41551-023-01056-8] [Medline: 37380750]
40. Cardoso SD, Da Silveira M, Pruski C. Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies. Knowledge Based Syst 2020;194:105508. [doi: 10.1016/j.knosys.2020.105508]
41. de la Torre J, Puig D, Valls A. Weighted kappa loss function for multi-class classification of ordinal data in deep learning. Pattern Recognit Lett 2018;105:144-154. [doi: 10.1016/j.patrec.2017.05.018]
42. Huang Y, Talwar A, Chatterjee S, Aparasu RR. Application of machine learning in predicting hospital readmissions: a scoping review of the literature. BMC Med Res Methodol 2021;21(1):96 [FREE Full text] [doi: 10.1186/s12874-021-01284-z] [Medline: 33952192]
43. Duncan R, Stewart K, Scheuermeyer F, Abu-Laban R, Ho K, Lavallee D, et al. Concordance between Virtual Physician advice and subsequent health service utilization for callers to a nurse-managed provincial health information telephone service: a time-to-event analysis. BMC Health Serv Res 2023. [doi: 10.21203/rs.3.rs-2733581/v1]
44. Blank L, Coster J, O'Cathain A, Knowles E, Tosh J, Turner J, et al. The appropriateness of, and compliance with, telephone triage decisions: a systematic review and narrative synthesis. J Adv Nurs 2012;68(12):2610-2621. [doi: 10.1111/j.1365-2648.2012.06052.x] [Medline: 22676805]
45. Tran DT, Gibson A, Randall D, Havard A, Byrne M, Robinson M, et al. Compliance with telephone triage advice among adults aged 45 years and older: an Australian data linkage study. BMC Health Serv Res 2017;17(1):512 [FREE Full text] [doi: 10.1186/s12913-017-2458-y] [Medline: 28764695]
46. Khader F, Kather JN, Müller-Franzes G, Wang T, Han T, Tayebi Arasteh S, et al. Medical transformer for multimodal survival prediction in intensive care: integration of imaging and non-imaging data. Sci Rep 2023 Jul 01;13(1):10666. [doi: 10.1038/s41598-023-37835-1] [Medline: 37393383]
47. Chen Z, Siltala-Li L, Lassila M, Malo P, Vilkkumaa E, Saaresranta T, et al. Predicting visit cost of obstructive sleep apnea using electronic healthcare records with transformer. IEEE J Transl Eng Health Med 2023;11:306-317 [FREE Full text] [doi: 10.1109/JTEHM.2023.3276943] [Medline: 37275471]

## Abbreviations

**AUC:** area under the curve
**ED:** emergency department
**K-NN:** K-nearest neighbors
**ML:** machine learning
**NB:** naïve Bayes
**NLP:** natural language processing
**NN:** neural network
**RF:** random forest
**SVM:** support vector machine
**XGBoost:** extreme gradient boosting

*Edited by T de Azevedo Cardoso; submitted 24.01.24; peer-reviewed by T Aslanidis, L Yperzeele, O Salman; comments to author 23.04.24; revised version received 13.05.24; accepted 18.07.24; published 11.09.24.*

*Please cite as:*
*Raff D, Stewart K, Yang MC, Shang J, Cressman S, Tam R, Wong J, Tammemägi MC, Ho K*
*Improving Triage Accuracy in Prehospital Emergency Telemedicine: Scoping Review of Machine Learning–Enhanced Approaches*
*Interact J Med Res 2024;13:e56729*
*URL: https://www.i-jmr.org/2024/1/e56729*
*doi:10.2196/56729*
*PMID:*

Review

# Assessing the Evidence for Nonobstetric Risk Factors for Deformational Plagiocephaly: Systematic Review and Meta-Analysis

Christopher Robert Timothy Hillyar[1], MBBS, BSc (Hons), MSc (Oxon), DPhil (Oxon), PGCert; Natalie Bishop[2], MBBS, iBSc; Anjan Nibber[1], BMBCh (Oxon), BSc (Hons), DPhil (Oxon); Frances Jean Bell-Davies[3], BA, BMBCh (Oxon); Juling Ong[4], MBBS, FRCS (Plast)

[1]Oxford Medical School, Green Templeton College, University of Oxford, Oxford, United Kingdom

[2]UCL Medical School, University College London, London, United Kingdom

[3]Department of Paediatrics, Wexham Park Hospital, Frimley Health NHS Foundation Trust, Wexham, United Kingdom

[4]Craniofacial Unit, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom

**Corresponding Author:**
Christopher Robert Timothy Hillyar, MBBS, BSc (Hons), MSc (Oxon), DPhil (Oxon), PGCert
Oxford Medical School
Green Templeton College
University of Oxford
43 Woodstock Rd
Oxford, OX2 6HG
United Kingdom
Phone: 44 01865 274770
Email: christopher.hillyar@gtc.ox.ac.uk

## Abstract

**Background:** Plagiocephaly is defined as an asymmetrical distortion of the skull, resulting in an oblique trapezoid or parallelogram head shape. Deformational plagiocephaly (DP) is caused by forces acting on one side of the back of the head, distorting normal skull symmetry.

**Objective:** The aims of this systematic review and meta-analysis were to critically assess the evidence for nonobstetric risk factors for DP and to make evidence-based recommendations for reducing the prevalence of DP.

**Methods:** The selection criterion was studies reporting risk factors for DP. Case reviews, case series, expert opinions, and systematic reviews were excluded. PubMed and Web of Science were searched from August 21, 2010, to August 21, 2022. Publication bias was assessed using funnel plots. Meta-analyses were presented using forest plots.

**Results:** A total of 19 studies (cohort studies: n=13, 68%; case-control studies: n=5, 26%; and cross-sectional studies: n=1, 5%) with a total of 14,808 participants were included. Of the 43 investigated potential nonobstetric factors, 16 (37%) were associated with DP. Of these 16 factors, 12 (75%) had odds ratios (ORs) with 95% CIs not crossing 1: insufficient vitamin D intake (OR 7.15, 95% CI 3.77-13.54), head position preference (OR 4.75, 95% CI 3.36-6.73), bottle-only feeding (OR 4.65, 95% CI 2.70-8.00), reduced tummy time (OR 3.51, 95% CI 1.71-7.21), sleeping position (OR 3.12, 95% CI 2.21-4.39), fewer motor milestones reached by the age of 6 months (OR 2.56, 95% CI 1.66-3.96), obesity (OR 2.45, 95% CI 1.02-5.90), maternal education level (OR 1.66, 95% CI 1.17-2.37), male sex (OR 1.51, 95% CI 1.07-2.12), formula feeding (OR 1.51, 95% CI 1.00-2.27), head circumference (OR 1.22, 95% CI 1.06-1.40), and mechanical ventilation (OR 1.10, 95% CI 1.00-1.14). No evidence of publication bias was detected.

**Conclusions:** This study provides a comprehensive assessment of the nonobstetric factors associated with DP and presents 11 evidence-based recommendations for reducing its prevalence. The primary limitation is that only publication bias was assessed.

**Trial Registration:** PROSPERO CRD42020204979; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020204979

***(Interact J Med Res 2024;13:e55695)*** doi:10.2196/55695

**KEYWORDS**

deformational plagiocephaly; plagiocephaly; flat head syndrome; back to sleep; meta-analysis; systematic review; meta-analyses; systematic reviews; vitamin D; vit D; head position preference; head position; head positioning; bottle feeding; tummy time; sleeping position; motor milestones; obesity; maternal education level; male sex; formula feeding; macrocephaly; head circumference; mechanical ventilation; pediatric; pediatrics; paediatric; paediatrics; infant; infants; infancy; baby; babies; neonate; neonates; neonatal; toddler; toddlers; child; children

## Introduction

### Background

Plagiocephaly is defined as an asymmetrical distortion of the skull resulting in an oblique trapezoid or parallelogram head shape when viewed from the vertex position in the axial plane [1]. The severity of skull asymmetry can range from minimal focal flattening on one side of the cranial vault to severe deformation affecting the entire cranial vault, skull base, and facial skeleton. Plagiocephaly arises via 2 main mechanisms: premature fusion of ≥1 of the cranial sutures (craniosynostosis) or external mechanical forces acting on the cranial vault, which results in a distortion of the normally symmetric craniofacial skeleton (deformational plagiocephaly [DP]).

In craniosynostotic plagiocephaly involving any of the paired coronal or lambdoid sutures, the restriction of skull vault growth occurs perpendicular to the fused suture (Virchow's law) [2,3]. Isolated craniosynostosis involving premature fusion of a single coronal suture results in an anterior plagiocephaly with brow retrusion on the affected side; similarly, craniosynostosis of a single lambdoid suture will restrict posterior cranial growth on the same side. The asymmetry is often accentuated as the remaining unfused sutures expand to enable accommodation of the rapidly growing infant brain. The majority of patients with craniosynostosis do not have an identifiable genetic cause, but this proportion is increased in patients with >1 suture involved [2,3].

Alternatively—and far more commonly—plagiocephaly is caused by deformational forces acting on one side of the back of the head, which distorts the normal symmetry of the skull in the absence of skull growth restriction due to craniosynostosis [4]. This deformity is characterized by a parallelogram-type deformity. This appears clinically as mild to severe occipital flattening, with or without ipsilateral anterior shift of the ear and orbital involvement [5]. The flattening of the posterior neurocranium, resulting from the external forces applied to the head, has led to the condition also being referred to as "flat head syndrome" [6]. The importance of external forces can be seen in the close relationship between DP and sleeping position [7], among other factors that may influence external head forces [8-12].

### Objectives

Since the 1980s, "back to sleep" campaigns have successfully publicized the benefits of supine sleeping for reducing the risk of sleep-related death, including sudden infant death syndrome [13]. Although these campaigns reduced the incidence of sudden infant death syndrome by 40%, an undesirable consequence has been an increase in the referrals of cases of DP, leading to more referrals to specialist centers [14,15]. The majority of cases of DP will resolve without intervention, and surgical treatment is not required [16]. However, in a subset of children, DP persists, even into teenage years [17]. Although physiotherapy and helmet therapy may play a role in improving head shape and limiting other long-term effects [18,19], understanding the factors that increase the risk of DP may help to prevent DP from developing. The aims of this systematic review and meta-analysis were to critically assess the evidence for risk factors for DP and to make evidence-based recommendations for reducing the prevalence of DP. This study was previously presented as a meeting abstract at the Royal College of Paediatrics and Child Health Conference on June 15, 2021.

## Methods

The study protocol, analysis, and reporting were conducted in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [20-22]. The protocol was registered with PROSPERO (CRD42020204979).

### Search Strategy

A search of PubMed and Web of Science was performed covering the period from August 21, 2010, to August 21, 2022 (this included an initial search from August 21, 2010, to August 20, 2020, and an update search from August 21, 2010, to August 20, 2022, to ensure that the list of included studies was as up to date as possible). The combination of PubMed and Web of Science provides >97.5% coverage of published literature [23,24]. To balance comprehensive coverage with a pragmatic approach to ensure that the study was completed with limited available resources, additional databases were not searched, and hand searching and gray literature searches were also not performed. The search terms included "plagiocephaly" AND "risk factor."

### Study Eligibility

Study titles and abstracts were screened and assessed for relevance by a single reviewer (CRTH, NB, or AN). Specifically, original studies were included if they assessed risk factors for DP. Studies in a non-English language, those involving nonhuman subjects, and low-quality or nonoriginal studies (meeting abstracts, reviews, case series, case reports, and editorials) were excluded. Duplicates were identified by assessing study titles and removed manually by CRTH. Full-text review was performed after screening by a single reviewer (CRTH, NB, or AN). Studies not reporting nonobstetric risk factors for DP were excluded. Studies reporting preventive measures such as tummy time were included. However, studies involving treatments such as physical therapy and helmet therapy were not included. As screening was performed by a single reviewer, there were no discrepancies to be resolved.

### Data Extraction and Reporting

Data extraction and reporting followed the PRISMA guidelines. However, due to resource limitations, data from eligible full-text articles were extracted by a single reviewer (CRTH, NB, or AN). The main outcomes of interest were odds ratios (ORs) and risk ratios (RRs) with 95% CIs, or significant associations, for risk factors for DP (or biomarkers for DP; eg, oblique diameter difference index). These outcomes were split into factors that were associated with DP and those that were not. Other items extracted from the eligible studies included authors, country of study, funding source, study design, study aims, total participants, percentage female, population assessed, age at baseline, and selection criteria.

### Meta-Analysis

A meta-analysis of ORs and RRs for factors associated with DP was performed. The inconsistency index ($I^2$) and a Q statistic for chi-square significance for specific *df* were calculated to assess interstudy heterogeneity. Both fixed effects and random effects were reported. *P* values for 95% CIs were calculated. Excel (Microsoft Corp) and Prism (GraphPad Software) were used for statistical analysis.

### Funnel Plots

Publication bias was assessed using funnel plots of ORs and RRs for DP risk factors against study precision (1/SE). The Egger test for asymmetry was conducted using linear regression, with $P<.05$ indicating publication bias [25].

## Results

### Overview

The searches of PubMed and Web of Science yielded 159 articles; after removing 35 (22%) duplicates, 124 (78%) articles were screened based on abstract content. Of these 124 articles, 77 (62.1%) were not relevant. Full-text screening of the remaining 47 articles resulted in the exclusion of 20 (43%) irrelevant articles (these did not report nonobstetric risk factors for DP), 3 (6%) non-English articles, and 5 (11%) articles that were not accessible. Thus, of the initial 159 articles, 19 (11.9%) were eligible for inclusion in this study (Figure 1). The characteristics of the included studies are presented in Table 1.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for study eligibility and inclusion.

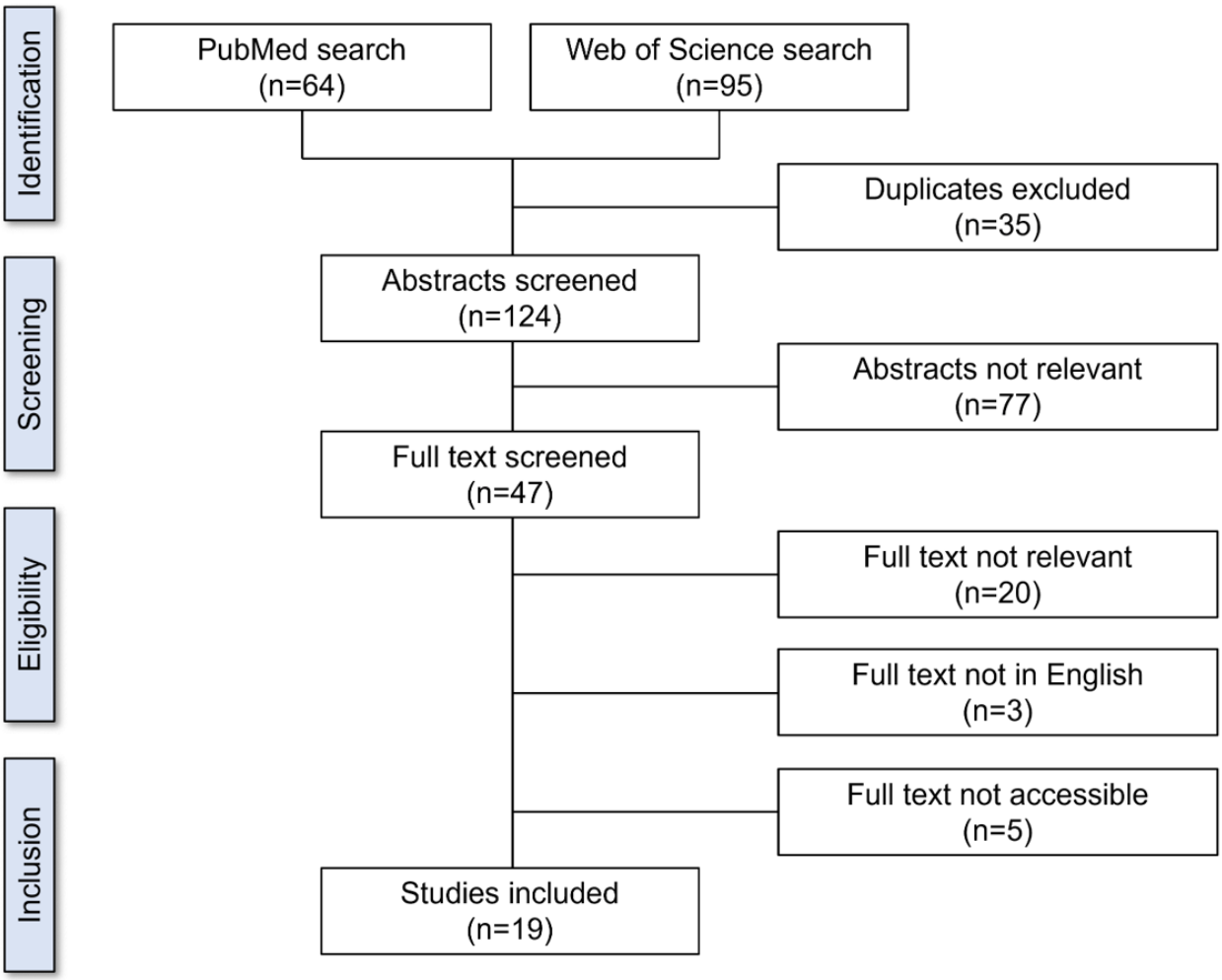


**Table 1.** Characteristics of the included studies.

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Aarnivala et al [26]; Finland | Alma and KA Snellman Foundation and The Foundation for Pediatric Research, Finland | Retrospective cohort study | To assess risk factors for the development of DP and torticollis | 155 (51) | Healthy newborns | <72 h from birth | Healthy newborns born between February 2012 and September 2013; newborns were excluded if they had chromosomal anomalies, cleft lip or palate, or craniosynostosis | ORs[b] and RRs[c] for risk factors for DP | NR[d] | Risk of developing DP was not associated with torticollis |
| Aarnivala et al [27]; Finland | Alma and KA Snellman Foundation | Prospective cohort study | To assess risk factors for cranial deformation by measuring cranial asymmetry from age 3 to 12 mo using 3D stereophotogrammetry | 99 (47) | Healthy newborns | Recruited at birth | Healthy newborns included at birth residing within 20 minutes' driving distance from Oulu University Hospital; infants with craniosynostosis, cleft lip and palate, or syndromic features were excluded | ORs and RRs for risk factors for DP | DP at age 6 mo associated with position preference and imbalance in head rotation at age 3 mo; at age 6 mo, DP associated with reaching fewer motor milestones at age 6 mo; at age 12 mo, DP associated with position preference at age 3 mo and imbalance in head rotation at age 3 mo and 6 mo; at age 12 mo, DP associated with fewer motor milestones and spending more time supine on the floor at age 6 mo; position preference at age 3 mo associated with DP at age 12 mo; at age 6 mo, DP associated with position preference at age 3 mo and fewer motor milestones reached at age 6 mo; at age 12 mo, DP associated with position preference at age 3 mo | Position preference at age 6 mo was not associated with DP at age 12 mo; there was no association between DP at age 12 mo and imbalance in head rotation at age 3 mo; no association with DP was also observed for primary sleeping position, time spent in carrier, bouncers, or car seats, time spent prone on the floor, pacifier use, illness history (acute otitis media and conditions requiring prolonged hospitalization), and duration of full breastfeeding |

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Ballardini et al [28]; Italy | Pediatric department of the University of Ferrara in Ferrara, Italy | Prospective cohort study | To assess prevalence and risk factors for DP in full-term infants | 283 (53) | Healthy newborns | All infants: mean 11.6 (range 9.4-12.9) wk; infants with DP: mean 11.7 (range 10.0-12.8) wk; infants without DP: mean 11.6 (range 9.4-12.9) wk | Healthy infants born at term presenting at a public immunization clinic in Ferrara at age 8 to 12 wk were included; infants affected by craniosynostosis, malformations, or neurological diseases or those admitted to the NICU[e] were excluded | ORs and RRs for risk factors for DP | Maternal and infant risk factors associated with DP included maternal age, supine sleeping position, and head position preference | Maternal and infant risk factors not associated with DP included infant sex, maternal origin, maternal education level, breast milk feeding, changing crib end, tummy time, and instruction about tummy time |
| Ifflaender et al [29]; Germany | The Else Kröner-Fresenius Foundation | Cross-sectional study | To assess head shape to determine the prevalence of symmetrical and asymmetrical head deformities and identify possible risk factors | 195 (48) | Healthy newborns | Mean postmenstrual age: 38.4 (SD 0.9) wk | Preterm infants discharged from an intermediate care unit of a tertiary neonatal clinic in Dresden from April 2011 to January 2013 were included; all infants were included that were present on the ward at the time of measurement; those with peripheral cannula at the scalp or requiring supplemental oxygen, were excluded | ORs and RRs for risk factors for DP | Cranial vault asymmetry index ([diagonal A – diagonal B and diagonal A] × 100, where diagonal A >diagonal B) at term-equivalent age was higher (4.1%, IQR 1.9%-6.5%) in very preterm infants compared to late preterm infants and term infants; moderate or severe DP at term-equivalent age was associated with intracranial hemorrhage in preterm infants; duration of total respiratory support was higher in cases of DP compared to controls; duration of continuous positive airway pressure therapy was longer in cases of DP than in controls | No association with DP was observed for sex, bronchopulmonary dysplasia, necrotizing enterocolitis, and duration of intermittent mandatory ventilation |

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Kim et al [30]; South Korea | NR | Case-control study | To determine risk factors for DP and its severity, including the association between DP and infant obesity as defined by BMI | 135 (40) | Infants with cranial deformation on physical examination, with age- and sex-matched controls randomly selected from the national health screening system for infants and children (confirmed to have no cranial deformation on physical examination) | 2-12 mo | Infants aged 2-12 mo at the time of diagnosis of DP were included; infants with neuroimaging results positive for craniosynostosis were excluded; infants with insufficient clinical data, congenital muscular torticollis, congenital anomalies (eg, craniofacial or chromosomal anomalies), or birth injury were also excluded; age- and sex-matched controls (confirmed to have no cranial deformation on physical examination) were randomly selected from the national health screening system | ORs and RRs for risk factors for DP | Factors associated with DP included bottle-only feeding, reduced tummy time when awake, delayed motor development, and obesity at diagnosis | Factors not associated with DP included; infant male sex; maternal age; macrocephaly at birth; and macrocephaly, underweight, or overweight at the time of diagnosis |
| Leung et al [31]; Australia | NR | Prospective cohort study | To assess association between DP and head orientation or head strength | 94 (59) | Healthy newborns | Mean 21.40 (SD 2.29) d | Healthy newborns, born at full term (37-42 wk) between June 2011 and July 2013; infants were excluded for reasons related to low birth weight (<2500 g), congenital muscular torticollis, craniosynostosis, neurological insult, or other medical or orthopedic conditions | Correlation between risk factor severity and DP severity | DP at age 9 wk was associated with asymmetrical head orientation duration at age 3 wk and 6 wk; DP at age 9 wk was significantly associated with asymmetrical head orientation strength at age 3 wk and 6 wk | No association between DP and latency to turn head |

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Maniglio et al [32]; Italy | NR | Prospective cohort study | To assess factors associated with DP to improve screening strategies to identify infants at risk of developing severe deformation | 4337 (NR) | Infants with DP and controls without DP | 6-12 wk | Infants born at the San Pietro Fatebenefratelli Hospital in Rome between January 2017 and September 2018 were included; infants with congenital deformations, born before 24 wk gestation, and infants who needed long intensive care treatment were excluded | ORs for risk factors for DP | Maternal age was associated with DP | Male sex was not associated with DP |
| Mawji et al [33]; Canada | The Faculty of Graduate Studies at the University of Calgary in Calgary, Alberta, provided CAD $3000 (US $2670) for data collection | Prospective cohort study | To assess potential risk factors for DP in infants aged 7-12 wk in Calgary | 440 (40.7) | Healthy full-term infants | 7-12 wk | Healthy full-term infants (born at ≥37 wk gestation) ranging from age 7 to 12 wk who presented for immunization at a 2-mo well-child clinic in Calgary were included | ORs and RRs for risk factors for DP | Significant difference in incidence of DP in infants placed supine to sleep compared with sleep in other positions (including prone; side; or a combination of supine, prone, or side); DP also associated with head position preference (right and left), maternal education level, and supine sleep position | No association of DP with average length of time in Canada; infant feeding position; length of tummy time received; male infant sex; and mothers who had a language barrier |

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Nuysink et al [34]; Netherlands | NR | Prospective cohort study | To assess predictive factors for DP in infants at corrected age >6 mo | 120 (45.8) | Newborns admitted to the NICU at gestational age <30 wk or birthweight <1000 g | <1 wk | Eligible infants were born in or referred to the level III neonatal intensive care unit within 1 week of birth between January 2009 and October 2010; infants born at gestational age <30 wk or birth weight <1000 g who visited the neonatal follow-up clinic were included; infants diagnosed with a disease or dysfunction leading to symptomatic asymmetry, such as a central nervous system disorder or congenital malformation, were excluded | ORs for risk factors for DP | Association between DP and mechanical ventilation and chronic lung disease grade II in the neonatal period | NR |
| Launonen et al [35]; Finland | The University of Oulu Scholarship Foundation, the Orthodontic Section of the Finnish Dental Association Apollonia, the Emil Aaltonen Foundation, the Alma, and KA Snellman Foundation, the Finnish Medical Foundation, and the Foundation for Pediatric Research in Finland | Case-control study | To use 3D stereophotogrammetry to assess cranial growth, molding, and incidence of DP in preterm children compared to term-born children | 68 (32) | Healthy newborns | Preterm (mean gestational age): 32.7 wk; term (mean gestational age): 40.0 wk | Infants were considered eligible if they had no cheilopalatoschisis, craniosynostosis, or dysmorphic features and if they resided within Oulu region, Finland; all participants were born between 2012 and 2015 at Oulu University Hospital; the control group was randomly selected from a previously collected noninterventional cohort by computer-based random selection and matched for sex | OCLR[f] mean difference | NR | No difference in head shape (OCLR) between preterm and full-term children or between sexes |

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Pogliani et al [36]; Italy | NR | Retrospective cohort study | To assess risk factors for DP at birth | 413 (50) | Healthy newborns | <72 h from birth | Neonates at gestational age >33 wk born between May 2011 and January 2012; newborns with extreme low birth weight or presenting with major congenital malformations needing NICU transfer were excluded; children of mothers with a documented TORCH[g] infection were also excluded | ORs and RRs for risk factors for DP | Association between DP and male sex | NR |
| Roberts et al [37]; United Kingdom | NR | Retrospective cohort study | To test the hypothesis that ventriculoperitoneal shunt insertion significantly increases contralateral DP | 339 (42) | Children aged 0-16 y with ventriculoperitoneal shunts | NR | Children aged 0-16 y with at least 1 follow-up scan from the surgical database at the pediatric neurosurgery department of Birmingham Children's Hospital between 2006 and 2013; children without imaging were excluded | ORs and RRs for risk factors for DP | DP was associated with occipital shunt placement; a statistically significant difference in the probability of becoming plagiocephalic between neonates and children aged 12-16 y was observed; boys were more likely to develop shunt-associated plagiocephaly than girls | No difference in DP between neonates and infants, neonates and children aged 3-5 y, neonates and children aged 1-3 y, and neonates and children aged 5-12 y |
| Sheu et al [38]; United States | Cooperative agreement from the Centers for Disease Control and Prevention (U01DD000494) and Title V Maternal and Child Health Block Grant funds from the Health Resources and Services Administration | Retrospective cohort study | To assess factors that may explain a 9-fold increase in plagiocephaly in Texas from 1999 to 2007 | 6295 (38) | Infants with DP | NR | Cases identified using the Texas Birth Defects Registry with a definite diagnosis of DP (British Paediatric Association code 754.050), born between January 1, 1999, and December 31, 2007; cases of plagiocephaly with craniosynostosis were excluded | Mean difference or percentage difference for risk factors for DP | Lower maternal education level was associated with DP | No association of DP with maternal age or race and ethnicity, infant sex; mean age at DP diagnosis did not significantly change over time |

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Solani et al [39]; Iran | Grant funding from the Vice-Chancellor for Research, Kashan University of Medical Sciences, Kashan, Iran | Case-control study | To determine the risk factors of positional plagiocephaly in healthy infants | 300 (NR) | Healthy Iranian infants | 8-12 wk | Healthy full-term (gestational age >37 wk) infants aged 8-12 wk who were referred to the pediatric neurology clinic at Shahid Beheshti Hospital in Kashan, affiliated with Kashan University of Medical Sciences, were included | ORs for risk factors for DP | Factors associated with DP included male sex, head circumference, and supine sleeping position | Firmness of headrest was not associated with DP |
| Tang et al [40]; United States | NR | Prospective cohort study | To assess the prevalence of DP in infants with NBPP[h] and spontaneous recovery from DP | 28 (50) | Full-term infants aged >3 mo and <1 y with NBPP | Mean 3 (SD 3) mo | Full-term infants aged >3 mo and <1 y with NBPP; infants with neurological or congenital comorbidities in addition to NBPP, helmet therapy for plagiocephaly, and surgical procedures related to NBPP were excluded; infants with craniosynostosis were also excluded | Mean difference or percentage difference for risk factors for DP | NR | Maternal age and race (Black or White) were not associated with DP |
| Valkama et al [41]; Finland | NR | Case-control study | To assess the prevalence of DP in children with DDH[i] | 120 (56) | Children with DDH with or without DP | Children with DDH: mean 8.0 (SD 1.4) y; matched controls: mean 7.9 (SD 1.3) y | Children with DDH from among newborn infants born at the Oulu University Hospital in Oulu, Finland, were included; preterm children and children with disabilities were excluded | ORs and RRs for risk factors for DP | 10.3% of the children with DDH and only 1.5% of the control children had DP | OCLR was equal between children with DDH and controls; no association between side of DDH and DP |

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Van Vlimmeren et al [42]; Netherlands | Grant from the Scientific Committee of The Royal Dutch Association for Physiotherapy (BU002/10) | Prospective cohort study | To assess skull shape in healthy newborns until age 5.5 y in children with position preference at 7 wk and those without and in children with position preference who received pediatric physical therapy intervention and those who did not | 380 (53) | Healthy newborns | <48 h from birth | Healthy newborns (gestational age ≥36 wk) born between December 2004 and September 2005 at Hospital Bernhoven; children with congenital muscular torticollis (Kaplan type 2 and 3), dysmorphism, or syndromes were excluded | Mean difference or percentage difference for risk factors for DP | Association between DP and position preference | No association between potential risk factors (nursing, feeding, sleeping, and playing positioning habits) at age 7 wk and skull deformity at age 24 mo and 5.5 y; a trend toward significance between time spent playing prone (tummy time) at age 7 wk and the ODDI[j] percentage at age 24 mo |
| Weernink et al [43]; Netherlands | ZonMw, the Netherlands Organization for Health Research and Development (170.992.501) | Case-control study | To assess the influence of adherence to recommendations for vitamin D supplement intake of 10 µg/d (400 IU) in the first months of life (child) on the occurrence of DP of the child at age 2-4 mo | 823 (46) | Infants with DP and those without | 2-4 mo | Children born between November 22, 2009, and June 9, 2010, with mild to severe DP from the Helmet Therapy Assessment in Deformed Skulls study; controls were included from a 2010 survey on infant milk feeding | ORs and RRs for risk factors for DP | Insufficient vitamin D supplement intake during early infancy was associated with DP; maternal sociodemographic factors associated with DP included mother's age and mother's education level; infant factors associated with DP included male sex, formula feeding, and milk formula consumption after birth | Maternal sociodemographic factors not significantly associated with DP included mother's country of birth (other than the Netherlands); infant factors not associated with DP included time child spent outdoors |

| Study; country | Funding source | Study design | Study aims | Participants, n (% female) | Population assessed | Age at baseline | Selection criteria | Outcomes | Factors associated with DP[a] | Factors not associated with DP |
|---|---|---|---|---|---|---|---|---|---|---|
| Van Cruchten et al [9]; Netherlands | NR | Cohort study | To assess the impact of risk factors on the type and severity of DP | 184 (30.4) | Infants seen at outpatient clinic with a parental concern for DP | 3-14 mo | Exclusion criteria consisted of children aged >14 mo or <3 mo and other forms of cranial deformation, such as cranial synostosis | Differences in ODDI for risk factors for DP | Negative correlation between age and ODDI; positive correlation between ODDI and position preference right and position preference left | No association between ODDI and developmental delay, family history of DP, sex, and torticollis |

[a]DP: deformational plagiocephaly.

[b]OR: odds ratio.

[c]RR: risk ratio.

[d]NR: not reported.

[e]NICU: neonatal intensive care unit.

[f]OCLR: oblique cranial length ratio.

[g]TORCH: toxoplasmosis, other (including infections such as syphilis, varicella-zoster, and parvovirus B19), rubella, cytomegalovirus, and herpes simplex virus.

[h]NBPP: neonatal brachial plexus palsy.

[i]DDH: developmental dysplasia of the hip.

[j]ODDI: oblique diameter difference index.

## Demographic Factors

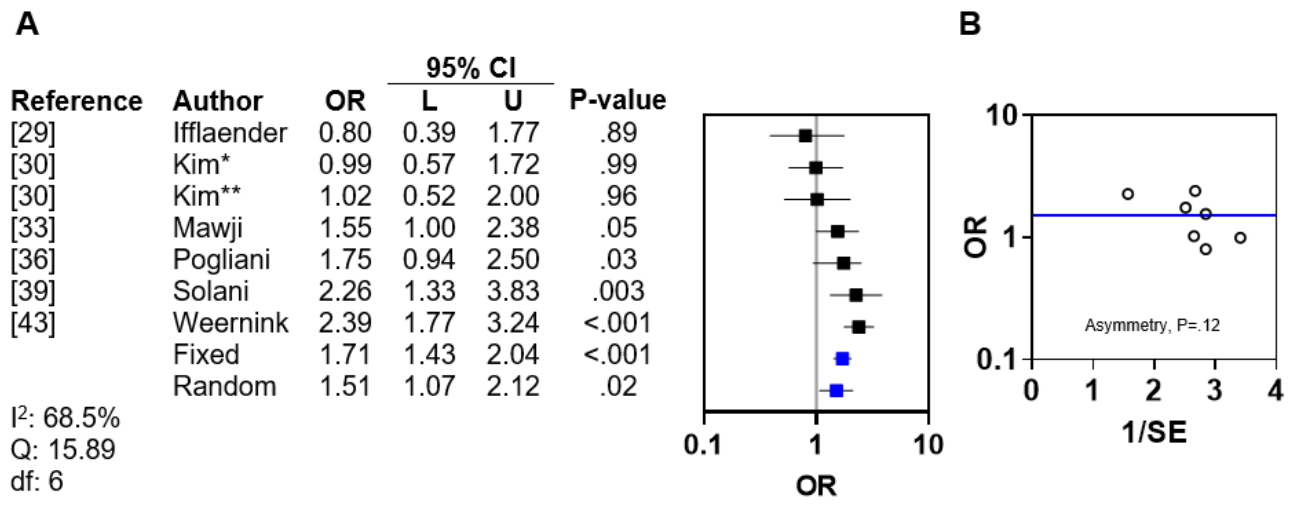
### *Age*

The study by van Cruchten et al [9] reported a negative correlation between age and oblique diameter difference index (a biomarker for DP). However, the study by Sheu et al [38] (which assessed factors associated with a 9-fold increase in plagiocephaly between 1999 and 2007 in Texas, United States) reported no association between age and DP. Finally, the study by Roberts et al [37], which assessed DP in children with ventriculoperitoneal shunts, reported that being aged 12 to 16 years at the time of shunt insertion was associated with DP.

### *Sex*

Of the 19 studies, 4 (21%) demonstrated an association between male sex and DP [33,36,39,44], 1 (5%) reported borderline association [37], and 7 (37%) reported no association between male sex and DP [9,28-30,32,35,38]. Of these 12 studies, 6 (50%) [29,30,33,36,39,43] reported ORs, and a meta-analysis of these ORs revealed interstudy heterogeneity ($I^2$=68.5%; Q statistic=15.89; *df*=6). The pooled fixed and random effects ORs for DP related to male sex were 1.71 (95% CI 1.43-2.04; *P*<.001) and 1.51 (95% CI 1.07-2.12; *P*=.02), respectively (Figure 2A [29,30,33,36,39,43]). Asymmetry analysis of the funnel plot excluded publication bias (*P*=.12; Figure 2B).

**Figure 2.** Meta-analysis and funnel plot for male sex. (A) Forest plot of odds ratios for deformational plagiocephaly related to male sex with fixed and random effects. (B) Funnel plot with linear regression test of asymmetry. The blue line indicates random effects. *Mild to moderate deformational plagiocephaly (DP; univariate analysis); **severe DP (univariate analysis).



### *Race*

Race was investigated by Tang et al [40], who included a cohort of children with brachial plexus palsy. In this group, race was not associated with DP that developed after brachial plexus injury.

## Developmental Factors

### *Developmental Delay*

Developmental delay was investigated as a risk factor for DP by van Cruchten et al [9], who concluded that developmental delay was not associated with DP.

### *Developmental Dysplasia of the Hip*

Developmental dysplasia of the hip (DDH) was investigated as a risk factor for DP by Valkama et al [41], who assessed the prevalence of DP in children with DDH. DDH was associated with DP compared to controls without DDH.

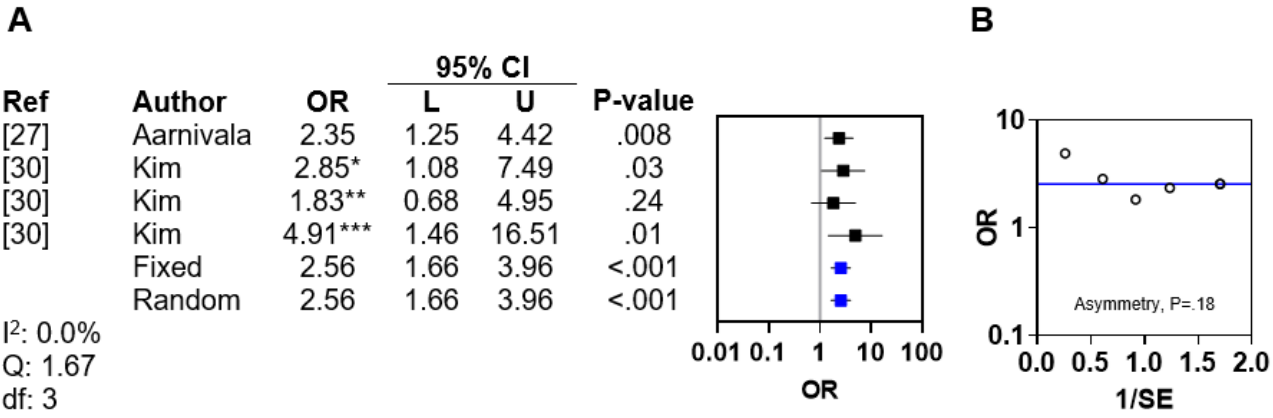
### *Head Circumference*

Head circumference was investigated by Solani et al [39], who concluded that it was associated with DP (OR 1.22, 95% CI 1.06-1.40).

### *Motor Milestones*

Reaching fewer motor milestones by age 6 months was investigated as a risk factor by 2 (11%) of the 19 included studies [27,30]. The first study reported that reaching fewer motor milestones by age 6 months was associated with DP (adjusted OR [aOR] 2.35, 95% CI 1.25-4.42) [27]. The second study found an association between delay in motor development and DP [30]. A meta-analysis of the ORs from these studies identified no interstudy heterogeneity ($I^2$=0%; Q statistic=1.67; *df*=3; Figure 3A [27,30]). The pooled fixed and random effects ORs for DP related to delayed motor milestones were both 2.56 (95% CI 1.66-3.96; *P*<.001). Asymmetry analysis of the funnel plot excluded publication bias (*P*=.18; Figure 3B).

**Figure 3.** Meta-analysis and funnel plot for reaching fewer motor milestones by age 6 months. (A) Forest plot of odds ratios for deformational plagiocephaly related to reaching fewer motor milestones by age 6 months with fixed and random effects. (B) Funnel plot with linear regression test of asymmetry. The blue line indicates fixed effects. *Adjusted OR (aOR) for deformational plagiocephaly (DP; multivariate analysis); **odds ratio (OR) for mild to moderate DP (only univariate analysis available); ***aOR for severe DP (multivariate analysis).




### *Overweight and Underweight*

Being overweight at diagnosis of DP was investigated as a risk factor for DP by Kim et al [30], who reported that being overweight at diagnosis was not associated with DP. The same study reported that being underweight at diagnosis was also not associated with DP.

## Dietary Factors

### *Bottle-Only Feeding*

Bottle-only feeding was investigated by Kim et al [30], who demonstrated that bottle-only feeding was associated with DP (aOR 4.65, 95% CI 2.70-8.00).

### *Breast Feeding*

The duration of exclusive breast feeding was investigated by Aarnivala et al [27], who demonstrated that the duration of exclusive breast feeding was not associated with DP.

### *Formula Feeding*

Formula feeding was investigated by Weernink et al [43], who reported that children who developed DP by age 2 to 4 months were more likely to be formula fed (aOR 1.51, 95% CI 1.00-2.27).

### *Vitamin D Intake*

Vitamin D intake in infants was investigated by Weernink et al [43], who reported that children who developed DP by age 2 to 4 months were more likely to have insufficient vitamin D intake (aOR 7.15, 95% CI 3.77-13.54).

## Maternal Factors

### *Maternal Age*

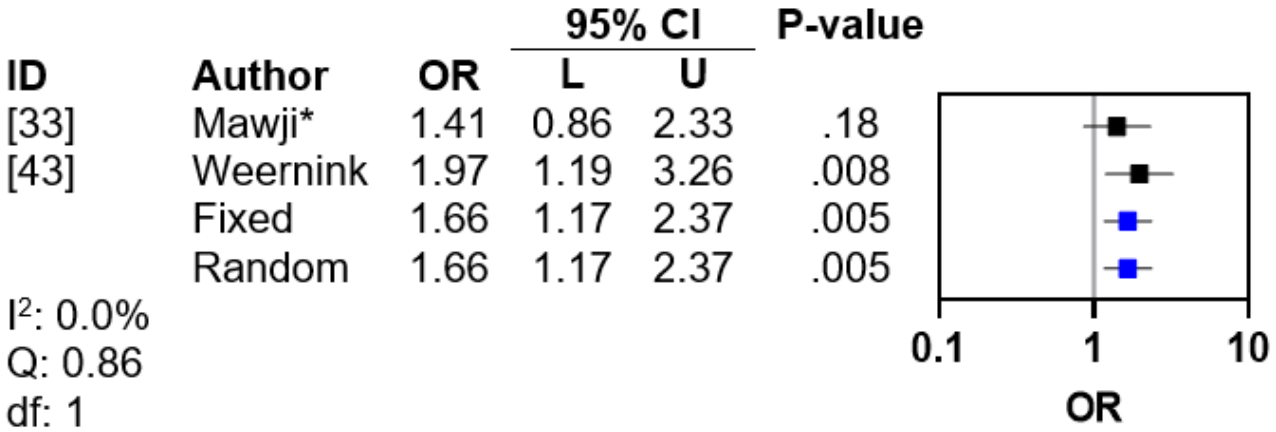
Of the 19 included studies, 3 (16%) demonstrated an association between maternal age and DP [28,32,43]. However, 3 (16%) of the 19 studies reported no association [30,38,40]. Only 1 (17%) of these 6 studies reported an OR indicating that increased maternal age was a protective factor for the development of DP (OR 0.94, 95% CI 0.91-0.97) [43].

### *Maternal Education Level*

Of the 19 included studies, 3 (16%) demonstrated an association between maternal education level and DP [33,38,43], and 1 (5%) reported no significant association [28]. Only 2 (67%) of

the 3 studies reported ORs [33,43]. Of these 2 studies, 1 (50%) indicated that low maternal education level was an adverse factor in the development of DP (aOR 1.97, 95% CI 1.19-3.26) [43], while 1 (50%) suggested that postsecondary education was not a protective factor (OR 0.71, 95% CI 0.43-1.16) [33]. The OR from the former was used in a meta-analysis with the inverse OR from the latter, which demonstrated no interstudy heterogeneity ($I^2$=0%; Q statistic=0.86; *df*=1; Figure 4 [33,43]). The pooled fixed effects OR for DP related to low education level was 1.66 (95% CI 1.17-2.37; *P*<.005).

**Figure 4.** Meta-analysis for maternal education level. *Reciprocal odds ratio (OR) for postsecondary education.




### *Maternal Language Barriers*

The study by Ballardini et al [28] investigated whether infants with mothers who experience a language barrier when receiving medical advice had a higher rate of DP. The study suggested that maternal language barrier was not associated with DP in the infant.

### *Maternal Race and Country of Origin*

Of the 19 included studies, 3 (16%) investigated whether DP was associated with maternal race and country of origin [28,38,43]. All 3 studies suggested that maternal race and country of origin was not associated with DP in the infant.

### *Length of Time in the Country of Study*

The study by Mawji et al [33] investigated whether the length of time spent by mothers in the country in which the study was conducted was associated with DP. The study reported that length of time in the country of study was not associated with DP.

### *Pacifier Use*

Pacifier use by mothers in healthy infants was investigated as a risk factor for DP by Aarnivala et al [27], who reported that pacifier use was not associated with DP.

### *Tummy Time Instructions*

The study by Ballardini et al [28] investigated whether DP was associated with mothers receiving instructions about tummy time. The study reported that receiving instructions about tummy time was not associated with DP.

## Medical and Surgical Factors

### *Bronchopulmonary Dysplasia*

Bronchopulmonary dysplasia was investigated as a risk factor for DP by Ifflaender et al [29], who included a cohort of infants born prematurely. The study reported that bronchopulmonary dysplasia was not associated with DP.

### *Chronic Lung Disease*

Chronic lung disease grade II was investigated by Launonen et al [34], who also assessed DP risk factors in infants born prematurely. The study reported that chronic lung disease grade II was associated with DP.

### *Family History of DP*

Family history of DP was investigated by van Cruchten et al [9]. The study suggested that family history of DP was not associated with DP.

### *History of Illness*

History of illness was investigated by Aarnivala et al [27]. The study suggested that history of illness was not associated with DP.

### *Intracranial Hemorrhage*

Intracranial hemorrhage was investigated by Ifflaender et al [29], who assessed DP risk factors in infants born prematurely. The study reported that intracranial hemorrhage was not associated with DP.

### *Macrocephaly*

Macrocephaly at birth was investigated by Kim et al [30]. The study suggested that macrocephaly at birth was not associated with DP. Macrocephaly at diagnosis of DP was also investigated, and it was found that this factor too was not associated with DP (OR 1.38, 95% CI 0.63-3.04) [30]. A lack of association was reported for subgroups with mild to moderate DP (OR 1.48, 95% CI 0.62-3.53) and severe DP (OR 1.19, 95% CI 0.04-3.58) [30].

### *Mechanical Ventilation*

Of the 19 included studies, 2 (11%) demonstrated an association between mechanical ventilation and the development of DP in preterm infants [29,34]. Of these 2 studies, 1 (50%) reported an OR of 1.10 (95% CI 1.00-1.14) for mechanical ventilation [34]. The other study also suggested that the duration of total

respiratory support (continuous positive airway pressure and intermittent mandatory ventilation) and the duration of continuous positive airway pressure alone were associated with DP, while intermittent mandatory ventilation alone was not associated with DP [29].

### *Necrotizing Enterocolitis*

Necrotizing enterocolitis was investigated by Ifflaender et al [29], who included a cohort of infants born prematurely. The study reported that necrotizing enterocolitis was not associated with DP.

### *Obesity*

Obesity at diagnosis of DP was investigated by Kim et al [30]. Obesity at diagnosis of DP was defined as BMI >97th percentile. The study concluded that obesity at diagnosis of DP was associated with DP (aOR 2.45, 95% CI 1.02-5.90). The study also suggested that obesity at diagnosis of DP was associated with severe DP (aOR 4.10, 95% CI 1.42-11.90) but was not associated with mild to moderate DP (aOR 2.29, 95% CI 0.86-6.05).

### *Occipital Shunt Placement*

Occipital shunt placement was investigated by Roberts et al [37], who concluded that occipital shunt placement was associated with DP compared to frontal shunt placement.

### *Torticollis*

Torticollis was investigated by 3 (16%) of the 19 included studies, all of which reported no association between torticollis and DP [9,26,38].

## Positional and Environmental Factors

### *Carriers, Bouncers, Car Seats, and Headrests*

Time spent in carriers, bouncers, and car seats was investigated by Aarnivala et al [27], who reported that time spent in carriers, bouncers, and car seats was not associated with DP. Firmness of headrest was investigated by Solani et al [39], who reported that firmness of headrest was not associated with DP (OR 1.31, 95% CI 0.72-2.37).

### *Change of Crib End*

Change of crib end (ie, alternating the infant's sleeping position by placing their head at different ends of the crib) was investigated as a risk factor for DP by 2 (11%) of the 19 included studies [33,42]. Both studies suggested that change of crib end was not associated with DP.

### *Feeding Position*

Feeding position was investigated as a risk factor for DP by 2 (11%) of the 19 included studies [33,42]; both demonstrated that feeding position was not associated with DP.

### *Latency in Head Turning*

Latency in head turning was investigated by Leung et al [31], who concluded that latency in head turning was not associated with DP.

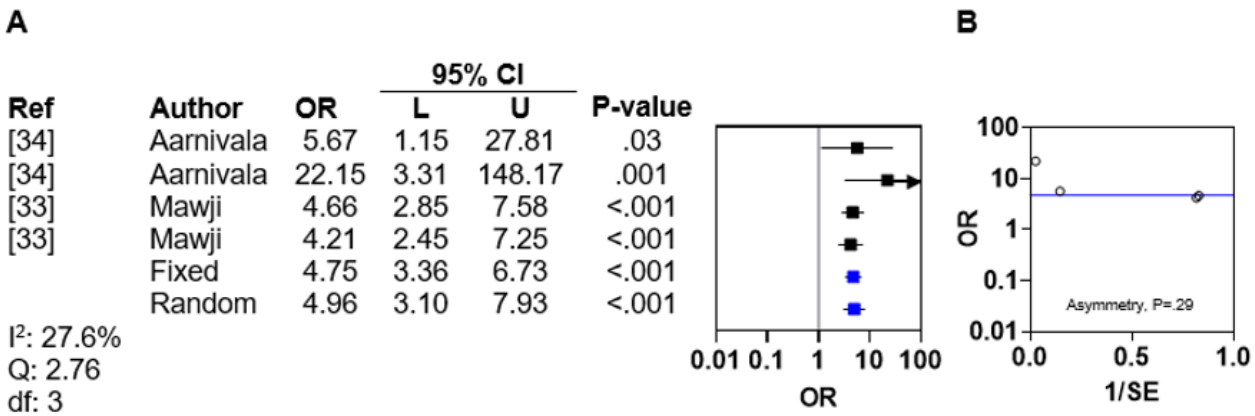
### *Playing Position*

Playing position was investigated by van Vlimmeren et al [42], who demonstrated that playing position was not associated with DP.

### *Head Position Preference*

Head position preference was investigated by 6 (32%) of the 19 included studies, all of which demonstrated an association between head position preference and DP [9,27,28,31,33,42]. Of these 6 studies, 2 (33%) [33,34] reported ORs, and a meta-analysis of these ORs revealed negligible interstudy heterogeneity ($I^2$=27.6%; Q statistic=2.76; *df*=3) (Figure 5A [33,34]). The pooled fixed and random effects ORs for DP related to head position preference were 4.75 (95% CI 3.36-6.73; *P*<.001) and 4.96 (95% CI 3.10-7.93; *P*<.001), respectively. Asymmetry analysis of the funnel plot excluded publication bias (*P*=.28; Figure 5B).

**Figure 5.** Meta-analysis and funnel plot for head position preference. (A) Forest plot of odds ratios for deformational plagiocephaly related to head position preference with fixed and random effects. (B) Funnel plot with linear regression test of asymmetry. The blue line indicates fixed effects.

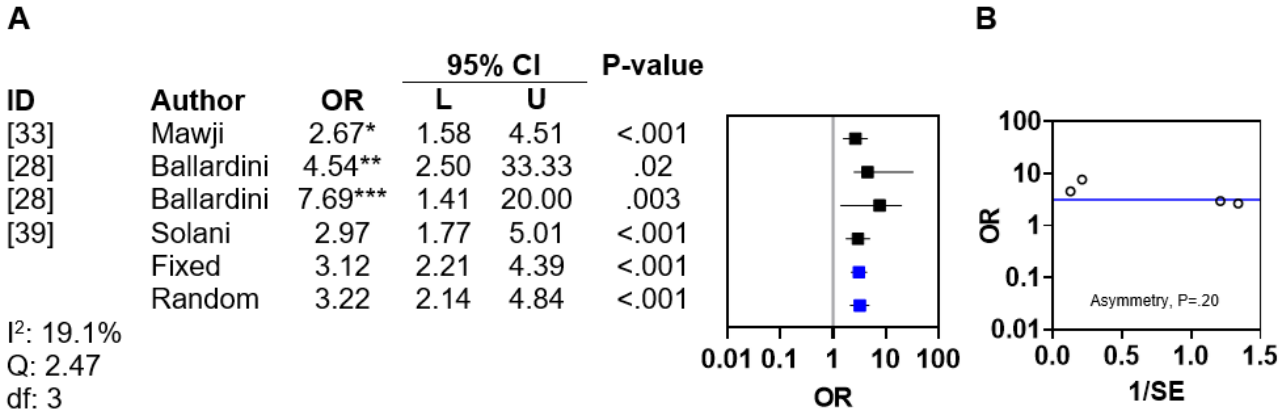



### *Sleeping Position*

Sleeping position (supine vs prone) was investigated by 3 (16%) of the 19 included studies [28,33,39]. Of these 3 studies, 2 (67%) demonstrated an association between supine sleeping and DP [33,39], while 1 (33%) reported a protective effect from prone sleeping (OR 0.13, 95% CI 0.03-0.40) or side sleeping (OR 0.22, 95% CI 0.05-0.71) [28]. The ORs from 2 (67%) [33,39] of the 3 studies were used with the inverse OR from the third study [28] in a meta-analysis, which demonstrated negligible

interstudy heterogeneity ($I^2$=19.1%; Q statistic=2.47; *df*=3; Figure 6A [28,33,39]). The pooled fixed and random effects ORs for DP related to sleeping position were 3.12 (95% CI 2.21-4.39; *P*<.001) and 3.22 (95% CI 2.14-4.84; *P*<.001), respectively. Asymmetry analysis of the funnel plot excluded publication bias (*P*=.20; Figure 6B).

**Figure 6.** Meta-analysis and funnel plot for supine sleeping position. (A) Forest plot of odds ratios for deformational plagiocephaly related to supine sleeping position with fixed and random effects. (B) Funnel plot with linear regression test of asymmetry. The blue line indicates fixed effects. *Supine versus prone; **reciprocal prone versus supine; ***reciprocal side versus supine.




### *Time Spent Prone on the Floor*

The study by Aarnivala et al [27] investigated time spent prone on the floor and reported no association with DP.

### *Tummy Time*

Tummy time was investigated by 4 (21%) of the 19 included studies [28,30,33,42]. Of these 4 studies, 1 (25%) reported an association between reduced tummy time when awake and DP (aOR 3.51, 95% CI 1.71-7.21) [30]. The remaining studies (3/4, 75%) reported that tummy time was not associated with DP but provided no ORs [28,33,42].

### *Time Spent Outdoors*

Time spent outdoors was investigated by Weernink et al [43], who reported no association with DP.

### *Time Spent Supine on the Floor*

Time spent supine on the floor was investigated as a risk factor for DP by Aarnivala et al [27], who reported that children with DP at age 12 months spent more time supine on the floor at age 6 months.

## Summary of Nonobstetric Factors Associated With DP

A summary of 16 nonobstetric factors associated with DP is presented in Table 2. ORs are provided for 12 (75%) of these 16 nonobstetric factors.

**Table 2.** Odds ratios (ORs) for factors associated with deformational plagiocephaly. ORs are reported from single studies unless marked with a superscript ("a" or "b") indicating pooled ORs from fixed or random effects meta-analyses.

| Factors | OR (95% CI) | References |
|---|---|---|
| Insufficient vitamin D intake | 7.15 (3.77-13.54) | [42] |
| Head position preference | 4.75[a] (3.36-6.73) | [33,34] |
| Bottle-only feeding | 4.65 (2.70-8.00) | [30] |
| Reduced tummy time | 3.51 (1.71-7.21) | [30] |
| Sleeping position | 3.12 (2.21-4.39) | [28,33,39] |
| Fewer motor milestones by age 6 mo | 2.56[a] (1.66-3.96) | [27,30] |
| Obesity (BMI >97th percentile) | 2.45 (1.02-5.90) | [30] |
| Maternal education level | 1.66[a] (1.17-2.37) | [33,43] |
| Male sex | 1.51[b] (1.07-2.12) | [29,30,33,36,43] |
| Formula feeding | 1.51 (1.00-2.27) | [43] |
| Head circumference | 1.22 (1.06-1.40) | [39] |
| Mechanical ventilation | 1.10 (1.00-1.14) | [34] |
| Chronic lung disease grade II | NR[c] | [34] |
| DDH[d] | NR | [41] |
| Maternal age | NR | [28,32,43] |
| Time spent supine | NR | [27] |

[a]OR from a fixed effects meta-analysis where interstudy heterogeneity was <50%.

[b]OR from a random effects meta-analysis where interstudy heterogeneity was ≥50%.

[c]NR: not reported.

[d]DDH: developmental dysplasia of the hip.

# Discussion

## Principal Findings

This study assessed evidence of association with DP for 43 nonobstetric factors (demographic factors: n=3, 7%; developmental factors: n=6, 14%; dietary factors: n=4, 9%; maternal factors: n=8, 19%; medical and surgical factors: n=11, 26%; and positional and environmental factors: n=11, 26%). Of these 43 factors, 16 (37%) were associated with DP (demographic factors [male sex]: n=1, 6%; developmental factors [DDH, head circumference, and delay in motor milestones]: n=3, 19%; dietary factors [bottle-only feeding, formula feeding, and vitamin D intake]: n=3, 19%; maternal factors [maternal age and maternal education level]: n=2, 12%; medical and surgical factors [chronic lung disease grade II, mechanical ventilation, and obesity at diagnosis of DP]: n=3, 19%; and positional and environmental factors [head position preference, sleeping position, reduced tummy time, and time spent supine on the floor]: n=4, 25%). With the notable exceptions of maternal age, mechanical ventilation, and tummy time, these associations were either supported by nonconflicting evidence or a meta-analysis that resolved conflicting evidence into an association. Of the 16 factors, 12 (75%) had ORs that ranged from 1.10 (mechanical ventilation) to 7.15 (insufficient vitamin D intake). Of the 5 factors assessed by meta-analysis (male sex, reaching fewer motor milestones by age 6 months, maternal education level, head position preference, and sleeping position), only 1 (20%; male sex) was associated with interstudy heterogeneity (≥50%). No evidence of publication bias was detected.

## Evidence-Based Recommendations

Strategies to reduce the prevalence of DP have included guidance about the infant's environment, positioning, and handling, with the goal of creating a nonrestrictive environment that promotes spontaneous and unhindered physical movement and symmetrical motor development [27,45].

On the basis of the evidence presented in this study, the following 11 recommendations, presented in order of importance, are offered with the aim of reducing the prevalence of DP.

First, to ensure bone health in infants, it is critical that vitamin D intake is adequate [46]. Vitamin D level should be assessed regularly during development, and dietary supplementation should be considered if vitamin D level is low.

Second, during the first months of life, babies develop a head position preference [47], and this preference is more often to the right [44,48]. The increased compressive forces on one side of the head for prolonged periods causes flattening on the side being compressed. It has been proposed that the position of the fetus in the later stages of pregnancy may, in part, be responsible for position preference [28,49]. However, there is also evidence that position preference can be modified by varying head

position during sleep to encourage equal distribution of pressure [50]. Other strategies to mitigate head position preference should be used, such as gently moving the infant's head to the unfavored side when asleep or physical therapy with or without kinesiological tape to reduce tightness in the neck muscles to facilitate easier neck turning [51,52].

Third, sufficient tummy time should be provided to strengthen the infant's head, neck, and arms and reduce time spent supine when awake.

Fourth, as infants tend to turn away from windows and toward the center of a room, it is recommended to alternate the infant's sleeping position by placing their head at different ends of the crib. This concept is supported by 2 studies suggesting no association between crib end change and DP [33,42]. However, it is important to state that the lack of association between crib end change and DP does not necessarily mean that changing the crib end will provide an effective treatment for established DP; it may merely help to reduce the risk of DP developing.

Fifth, although bottle-only feeding should not be discouraged, it is recommended to alternate feeding positions, using both the dominant and nondominant sides when holding the infant. Bottle feeding, as opposed to exclusively breast feeding, may be positively associated with obesity (which has also been identified as a risk factor for DP) [53]. This coassociation may, in part, explain the association of bottle-only feeding with DP.

Sixth, although formula feeding should also not be discouraged, it is recommended to alternate feeding positions. It has been suggested by Weernink et al [43] that this particular risk factor is associated with lower maternal education level. Similar to bottle feeding, formula feeding, as opposed to exclusively breast feeding, may be positively associated with obesity (itself a risk factor for DP) [53].

Seventh, if motor milestones are delayed, infants should be referred for specialist assessment by a pediatrician. Similarly, obesity (BMI >97th percentile) should be identified and managed by a specialist pediatrician.

Eighth, to mitigate the impact of low maternal education level, educational resources and face-to-face education outlining factors associated with DP, as well as recommendations to reduce their impact, should be provided to all families at the 6-week postnatal infant check-up, with particular emphasis on those with low educational levels. A lower maternal education level results in worse health outcomes in infancy and later life [54]; although improving the education level of future mothers should be a national priority, there are other factors associated with low education level that may be immediately modifiable. These include access to information and access to resources [55].

Ninth, Information should be provided to families that male infants are at higher risk of DP. Losee et al [49] suggested that male infants tend to have larger, less flexible heads at birth, which are more likely to become deformed by compressive forces in utero and in the birth canal. Families should be particularly encouraged to engage with strategies to mitigate DP in male infants.

Tenth, information should be provided to families that infants with greater head circumference are at higher risk of DP. Families should be particularly encouraged to engage with strategies to mitigate DP in these infants.

Eleventh, information should also be provided to families that infants who have had mechanical ventilation are at higher risk of DP. Families and care teams of infants requiring mechanical ventilation should be encouraged to engage with strategies to mitigate DP in these infants.

Alternative strategies for mitigating the impact of DP, such as helmet therapy, although costly for individuals and health care systems, may help avoid more invasive strategies [56]. It has been demonstrated that education provided by health care professionals, such as health visitors, midwives, and nurses, can successfully reduce the prevalence of DP [57]. Adopting these recommendations may lead to a reduction in the prevalence of DP.

### Strengths of This Study

The strengths of this study include a robust methodology. The protocol was registered with PROSPERO, and the reporting was in line with the PRISMA 2020 guidelines [22]. The databases searched provided >97.5% coverage [23]. Standardized data extraction was performed to minimize errors. Meta-analysis was performed where data allowed. Funnel plots suggested that publication bias did not impact the results; thus, a trim-and-fill analysis was not necessary to correct for asymmetry [58].

This study represents the most comprehensive analysis of nonobstetric factors associated with DP published to date. A systematic review by Bialocerkowski et al [59] only identified 5 factors associated with DP, of which 3 (60%) were nonobstetric factors (male sex, supine position, and neck problems) [59]. Another systematic review by De Bock et al [60] identified male sex, supine sleeping position, limited neck rotation, head position preference, lower level of activity, and reduced tummy time as the most important risk factors [60]. By contrast, our study provides evidence of association for 16 factors and confirms the association of male sex, supine sleeping position, head position preference, and reduced tummy time with DP.

A more recent systematic review by Inchingolo et al [61] provided only 2 recommendations to mitigate the impact of nonobstetric risk factors for DP: at least 30 minutes of tummy time and the use of a passive sleep curve mattress to improve harmonious skull growth. In addition, it did not include meta-analyses to support these recommendations. By contrast, our study makes 11 recommendations for nonobstetric factors to reduce the prevalence of DP. The systematic review by Inchingolo et al [61] also did not assess publication bias. Our review suggested that none of the recommendations were influenced by publication bias.

### Limitations of This Study

To ensure the completion of the study, additional databases were not searched, and hand searching and gray literature searches were also not performed. This inevitably limited the

comprehensiveness of this study, although no systematic review can claim to be truly comprehensive because this would necessitate continual inclusion of newly published studies. The search terms were designed to be specific yet pragmatic to ensure the completion of the work with limited resources; thus, some studies may have been overlooked. Due to resource limitations, it was not possible to record more detailed reasons for the exclusion of screened articles other than not reporting risk factors for DP. Resource constraints also prevented the screening process from being conducted by 2 independent reviewers. Instead, abstract and full-text screening was performed by a single reviewer. This impacts the reproducibility of the study due to the increased potential for errors in study selection when screening is undertaken by a single reviewer. That said, because the selection criteria were clearly set out, errors due to the application of these criteria during the screening process are likely to have had minimal impact on the study outcomes. In addition, it was not feasible, due to resource limitations, to contact study authors to attempt to collect raw data if a study did not report ORs for a risk factor. This may have limited the data that could have been used in meta-analysis. If contacting authors had been feasible, more factors could potentially have been analyzed quantitively to resolve discrepancies between the included studies. This represents a significant limitation of our study.

These factors may, in part, account for the apparent disparities reported in our results compared to other literature. These disparities include the apparent lack of association between developmental delay and DP [9], while demonstrating a significant association between reaching fewer motor milestones and DP [27,30]. A recent systematic review by Martinuik et al [62] included 19 studies that assessed the association between developmental delay and DP. Notwithstanding the fact that the authors included multiple studies that used the same study population more than once, a positive association between developmental delay and DP was reported in a majority of the studies [62]. The fact that our study did not demonstrate a similar association may, in part, be due to the inclusion of studies that assessed different populations, as well as methodological limitations in our search strategy that limited the identification of studies that may have met our eligibility criteria. Another study of a large primary care cohort of 77,108 patients has provided further evidence in support of an association between developmental delay and DP [63]. That said, the literature remains conflicting, with other studies unable to demonstrate an association between presence or degree of developmental delay and DP [64]. Commentators have highlighted the fact that most studies are retrospective and observational by design, and this limits conclusions about the correlative versus causative relationship between developmental delay and DP [65].

Another disparity reported in our study is that torticollis was not associated with DP [9,26,38]. A number of literature reviews have previously reported an association between congenital muscular torticollis and DP [8,59]. Although these studies were published a decade or so ago, it logically follows that if other positional factors, such as head position preference and sleeping position, have been found in more contemporaneous studies to be significantly associated with DP [28,33,34,39], then torticollis would also be expected to be found to be associated with DP. These conflicts reported in the literature may be due to the limitations of individual study design or differences between different study populations. Therefore, we recommend interpreting with caution the lack of association between torticollis and DP reported in our study, given the lack of a mechanistic explanation that reconciles this result with other positional factors that were found to be associated with DP.

In this review, we have included a study involving patients who developed plagiocephaly after ventriculoperitoneal shunt insertion [37]. Although the authors characterized the plagiocephaly as “positional,” implying that external forces had caused the skull deformity, an alternative hypothesis is that shunt-associated plagiocephaly is a different disease entity from DP [37]. This study has been included for comprehensiveness [37], but the results should be interpreted with caution alongside those of other studies.

Finally, another limitation includes the lack of a Grading of Recommendations Assessment, Development, and Evaluation (GRADE) assessment of evidence quality, which impacts the certainty of the conclusions and recommendations set out in this study.

## Conclusions

In summary, this study provides the most comprehensive meta-analytic assessment of nonobstetric factors associated with DP published to date. It offers 11 evidence-based recommendations that can be adopted by health care systems globally to reduce the prevalence of DP. Future research should focus on investigating factors for which the literature is conflicting but quantitative data are lacking to enable meta-analyses to be performed; for example, maternal age was the only factor reported to be protective against DP, but conflicting studies reported that there was no association without providing quantitative data. Thus, maternal age as a protective factor for DP should be investigated further to provide quantitative data for meta-analytic approaches to determine its protective effect. Further research should also address the nature of specific relationships between risk factors and DP; for example, both bottle-only feeding and formula feeding have been associated with DP. Studies should focus on understanding the nature of this relationship, that is, whether this relationship is due to mechanical forces associated with bottle-feeding or whether a reduction in, or lack of, breast milk intake or a lack of complimentary foods alongside breast milk results in nutritional differences that impact skull development [66]. As our study also highlighted an association between obesity and DP, research on other early-life or environmental exposures should be conducted to elucidate their effects on growth and development, particularly skull growth and head shape [67,68]. Finally, randomized controlled trials, although considered the gold standard study design for obtaining reliable evidence of an intervention’s effectiveness, should not be conducted for the interventions herein that relate to risk factors for DP development that have already been assessed through meta-analysis. However, randomized controlled trials could provide further evidence for an intervention—if ethically

appropriate—where the evidence for a particular risk factor is relatively weak, such as when only a single cohort study has provided the evidence for a risk factor's influence on the development of DP.

## Data Availability

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

## Authors' Contributions

CRTH developed the methodology, performed screening and data extraction, performed data analysis, interpreted the results, and wrote and edited the manuscript. NB performed screening and data extraction, interpreted the results, and wrote and edited the manuscript. AN performed screening and data extraction, interpreted the results, and edited the manuscript. FJBD interpreted the results and edited the manuscript. JO conceived of the study, interpreted the results, and edited the manuscript.

## Conflicts of Interest

CRTH and AN received consultancy fees from Advanced Orthomolecular Research in Calgary, Alberta, and their family members work for the aforementioned company, which, although not directly connected to this research, produces nutraceutical compounds, including vitamin D, which is part of the subject matter of this research. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
PRISMA 2020 checklist.
[PDF File (Adobe PDF File), 66 KB - ijmr_v13i1e55695_app1.pdf ]

## References


1. Ehret FW, Whelan MF, Ellenbogen RG, Cunningham ML, Gruss JS. Differential diagnosis of the trapezoid-shaped head. Cleft Palate Craniofac J 2004 Jan;41(1):13-19. [doi: 10.1597/02-053] [Medline: 14697075]
2. Delashaw JB, Persing JA, Broaddus WC, Jane JA. Cranial vault growth in craniosynostosis. J Neurosurg 1989 Feb;70(2):159-165. [doi: 10.3171/jns.1989.70.2.0159] [Medline: 2913214]
3. Johnson D, Wilkie AO. Craniosynostosis. Eur J Hum Genet 2011 Apr;19(4):369-376 [FREE Full text] [doi: 10.1038/ejhg.2010.235] [Medline: 21248745]
4. Jung BK, Yun IS. Diagnosis and treatment of positional plagiocephaly. Arch Craniofac Surg 2020 Apr;21(2):80-86 [FREE Full text] [doi: 10.7181/acfs.2020.00059] [Medline: 32380806]
5. Holowka MA, Reisner A, Giavedoni B, Lombardo JR, Coulter C. Plagiocephaly severity scale to aid in clinical treatment recommendations. J Craniofac Surg 2017 May;28(3):717-722. [doi: 10.1097/SCS.0000000000003520] [Medline: 28468155]
6. Kalra R, Walker ML. Posterior plagiocephaly. Childs Nerv Syst 2012 Sep;28(9):1389-1393. [doi: 10.1007/s00381-012-1784-y] [Medline: 22872253]
7. Marshall JM, Shahzad F. Safe sleep, plagiocephaly, and brachycephaly: assessment, risks, treatment, and when to refer. Pediatr Ann 2020 Oct 01;49(10):e440-e447. [doi: 10.3928/19382359-20200922-02] [Medline: 33034660]
8. Kuo AA, Tritasavit S, Graham Jr JM. Congenital muscular torticollis and positional plagiocephaly. Pediatr Rev 2014 Feb;35(2):79-87. [doi: 10.1542/pir.35-2-79] [Medline: 24488831]
9. van Cruchten C, Feijen MM, van der Hulst RR. Demographics of positional plagiocephaly and brachycephaly; risk factors and treatment. J Craniofac Surg 2021;32(8):2736-2740. [doi: 10.1097/SCS.0000000000007811] [Medline: 34231510]
10. van Vlimmeren LA, van der Graaf Y, Boere-Boonekamp MM, L'Hoir MP, Helders PJ, Engelbert RH. Risk factors for deformational plagiocephaly at birth and at 7 weeks of age: a prospective cohort study. Pediatrics 2007 Feb;119(2):e408-e418. [doi: 10.1542/peds.2006-2012] [Medline: 17272603]
11. Joganic JL, Lynch JM, Littlefield TR, Verrelli BC. Risk factors associated with deformational plagiocephaly. Pediatrics 2009 Dec;124(6):e1126-e1133. [doi: 10.1542/peds.2008-2969] [Medline: 19917588]
12. Oh AK, Hoy EA, Rogers GF. Predictors of severity in deformational plagiocephaly. J Craniofac Surg 2009 Mar;20 Suppl 1:685-689. [doi: 10.1097/SCS.0b013e318193d6e5] [Medline: 19190504]
13. de Luca F, Hinde A. Effectiveness of the 'Back-to-Sleep' campaigns among healthcare professionals in the past 20 years: a systematic review. BMJ Open 2016 Sep 30;6(9):e011435 [FREE Full text] [doi: 10.1136/bmjopen-2016-011435] [Medline: 27694485]
14. Turk AE, McCarthy JG, Thorne CH, Wisoff JH. The "back to sleep campaign" and deformational plagiocephaly: is there cause for concern? J Craniofac Surg 1996 Jan;7(1):12-18. [doi: 10.1097/00001665-199601000-00006] [Medline: 9086896]
15. Argenta LC, David LR, Wilson JA, Bell WO. An increase in infant cranial deformity with supine sleeping position. J Craniofac Surg 1996 Jan;7(1):5-11. [doi: 10.1097/00001665-199601000-00005] [Medline: 9086895]

16. Esparza J, Hinojosa J, García-Recuero I, Romance A, Pascual B, Martínez de Aragón A. Surgical treatment of isolated and syndromic craniosynostosis. Results and complications in 283 consecutive cases. Neurocirugia (Astur) 2008 Dec;19(6):509-529. [doi: 10.1016/s1130-1473(08)70201-x] [Medline: 19112545]
17. Roby BB, Finkelstein M, Tibesar RJ, Sidman JD. Prevalence of positional plagiocephaly in teens born after the "Back to Sleep" campaign. Otolaryngol Head Neck Surg 2012 May;146(5):823-828. [doi: 10.1177/0194599811434261] [Medline: 22241785]
18. van Wijk RM, van Vlimmeren LA, Groothuis-Oudshoorn CG, Van der Ploeg CP, Ijzerman MJ, Boere-Boonekamp MM. Helmet therapy in infants with positional skull deformation: randomised controlled trial. BMJ 2014 May 01;348:g2741 [FREE Full text] [doi: 10.1136/bmj.g2741] [Medline: 24784879]
19. Di Chiara A, La Rosa E, Ramieri V, Vellone V, Cascone P. Treatment of deformational plagiocephaly with physiotherapy. J Craniofac Surg 2019 Oct;30(7):2008-2013. [doi: 10.1097/SCS.0000000000005665] [Medline: 31232996]
20. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 2019 Oct 03;10:ED000142. [doi: 10.1002/14651858.ED000142] [Medline: 31643080]
21. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
23. Bramer WM, Giustini D, Kramer BM. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Syst Rev 2016 Mar 01;5(1):39 [FREE Full text] [doi: 10.1186/s13643-016-0215-7] [Medline: 26932789]
24. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008 Feb;22(2):338-342. [doi: 10.1096/fj.07-9492LSF] [Medline: 17884971]
25. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol 2001 Oct;54(10):1046-1055. [Medline: 11576817]
26. Aarnivala HE, Valkama AM, Pirttiniemi PM. Cranial shape, size and cervical motion in normal newborns. Early Hum Dev 2014 Aug;90(8):425-430. [doi: 10.1016/j.earlhumdev.2014.05.007] [Medline: 24951081]
27. Aarnivala H, Vuollo V, Harila V, Heikkinen T, Pirttiniemi P, Holmström L, et al. The course of positional cranial deformation from 3 to 12 months of age and associated risk factors: a follow-up with 3D imaging. Eur J Pediatr 2016 Dec;175(12):1893-1903. [doi: 10.1007/s00431-016-2773-z] [Medline: 27624627]
28. Ballardini E, Sisti M, Basaglia N, Benedetto M, Baldan A, Borgna-Pignatti C, et al. Prevalence and characteristics of positional plagiocephaly in healthy full-term infants at 8-12 weeks of life. Eur J Pediatr 2018 Oct;177(10):1547-1554. [doi: 10.1007/s00431-018-3212-0] [Medline: 30030600]
29. Ifflaender S, Rüdiger M, Konstantelos D, Wahls K, Burkhardt W. Prevalence of head deformities in preterm infants at term equivalent age. Early Hum Dev 2013 Dec;89(12):1041-1047. [doi: 10.1016/j.earlhumdev.2013.08.011] [Medline: 24016482]
30. Kim E, Kim KE, Jeon J, Sheen YH, Lee H, Yoon SY, et al. Delayed motor development and infant obesity as risk factors for severe deformational plagiocephaly: a matched case-control study. Front Pediatr 2020;8:582360 [FREE Full text] [doi: 10.3389/fped.2020.582360] [Medline: 33262962]
31. Leung AY, Mandrusiak A, Watter P, Gavranich J, Johnston LM. Clinical assessment of head orientation profile development and its relationship with positional plagiocephaly in healthy term infants - a prospective study. Early Hum Dev 2016 May;96:31-38. [doi: 10.1016/j.earlhumdev.2016.03.001] [Medline: 27031440]
32. Maniglio P, Noventa M, Tartaglia S, Petracca M, Bonito M, Ricciardi E, et al. The Obstetrician Gynecologist's role in the screening of infants at risk of severe plagiocephaly: prevalence and risk factors. Eur J Obstet Gynecol Reprod Biol 2022 May;272:37-42. [doi: 10.1016/j.ejogrb.2022.03.011] [Medline: 35279639]
33. Mawji A, Vollman AR, Fung T, Hatfield J, McNeil DA, Sauvé R. Risk factors for positional plagiocephaly and appropriate time frames for prevention messaging. Paediatr Child Health 2014 Oct;19(8):423-427 [FREE Full text] [doi: 10.1093/pch/19.8.423] [Medline: 25382999]
34. Nuysink J, Eijsermans MJ, van Haastert IC, Koopman-Esseboom C, Helders PJ, de Vries LS, et al. Clinical course of asymmetric motor performance and deformational plagiocephaly in very preterm infants. J Pediatr 2013 Sep;163(3):658-65.e1. [doi: 10.1016/j.jpeds.2013.04.015] [Medline: 23706356]
35. Launonen AM, Aarnivala H, Kyteas P, Vuollo V, Heikkinen T, Kau CH, et al. A 3D follow-up study of cranial asymmetry from early infancy to toddler age after preterm versus term birth. J Clin Med 2019 Oct 11;8(10):387-392 [FREE Full text] [doi: 10.3390/jcm8101665] [Medline: 31614700]
36. Pogliani L, Cerini C, Vivaldo T, Duca P, Zuccotti GV. Deformational plagiocephaly at birth: an observational study on the role of assisted reproductive technologies. J Matern Fetal Neonatal Med 2014 Feb;27(3):270-274. [doi: 10.3109/14767058.2013.814629] [Medline: 23773121]

37. Roberts SA, Symonds JD, Chawla R, Toman E, Bishop J, Solanki GA. Positional plagiocephaly following ventriculoperitoneal shunting in neonates and infancy-how serious is it? Childs Nerv Syst 2017 Feb;33(2):275-280 [FREE Full text] [doi: 10.1007/s00381-016-3275-z] [Medline: 27848003]
38. Sheu SU, Ethen MK, Scheuerle AE, Langlois PH. Investigation into an increase in plagiocephaly in Texas from 1999 to 2007. Arch Pediatr Adolesc Med 2011 Aug;165(8):708-713. [doi: 10.1001/archpediatrics.2011.42] [Medline: 21464373]
39. Solani B, Talebian Ardestani M, Boroumand H, Ostadmohammadi V, Hallajnejad M, Kashani Zade M, et al. Risk factors associated with positional plagiocephaly in healthy Iranian infants: a case-control study. Iran J Child Neurol 2022;16(2):85-92 [FREE Full text] [doi: 10.22037/ijcn.v16i1.28524] [Medline: 35497110]
40. Tang M, Gorbutt KA, Peethambaran A, Yang L, Nelson VS, Chang KW. High prevalence of cranial asymmetry exists in infants with neonatal brachial plexus palsy. J Pediatr Rehabil Med 2016 Nov 30;9(4):271-277. [doi: 10.3233/PRM-160396] [Medline: 27935564]
41. Valkama AM, Aarnivala HI, Sato K, Harila V, Heikkinen T, Pirttiniemi P. Plagiocephaly after neonatal developmental dysplasia of the hip at school age. J Clin Med 2019 Dec 19;9(1):21 [FREE Full text] [doi: 10.3390/jcm9010021] [Medline: 31861739]
42. van Vlimmeren LA, Engelbert RH, Pelsma M, Groenewoud HM, Boere-Boonekamp MM, der Sanden MW. The course of skull deformation from birth to 5 years of age: a prospective cohort study. Eur J Pediatr 2017 Jan;176(1):11-21 [FREE Full text] [doi: 10.1007/s00431-016-2800-0] [Medline: 27815732]
43. Weernink MG, van Wijk RM, Groothuis-Oudshoorn CG, Lanting CI, Grant CC, van Vlimmeren LA, et al. Insufficient vitamin D supplement use during pregnancy and early childhood: a risk factor for positional skull deformation. Matern Child Nutr 2016 Jan;12(1):177-188 [FREE Full text] [doi: 10.1111/mcn.12153] [Medline: 25382635]
44. Hopkins B, Lems YL, Van Wulfften Palthe T, Hoeksma J, Kardaun O, Butterworth G. Development of head position preference during early infancy: a longitudinal study in the daily life situation. Dev Psychobiol 1990 Jan;23(1):39-53. [doi: 10.1002/dev.420230105] [Medline: 2340956]
45. Cavalier A, Picot MC, Artiaga C, Mazurier E, Amilhau M, Froye E, et al. Prevention of deformational plagiocephaly in neonates. Early Hum Dev 2011 Aug;87(8):537-543. [doi: 10.1016/j.earlhumdev.2011.04.007] [Medline: 21664772]
46. Pawley N, Bishop NJ. Prenatal and infant predictors of bone health: the influence of vitamin D. Am J Clin Nutr 2004 Dec;80(6 Suppl):1748S-1751S [FREE Full text] [doi: 10.1093/ajcn/80.6.1748S] [Medline: 15585799]
47. Vles J, van Zutphen S, Hasaart T, Dassen W, Lodder J. Supine and prone head orientation preference in term infants. Brain Dev 1991;13(2):87-90. [doi: 10.1016/s0387-7604(12)80112-4] [Medline: 1892225]
48. Rönnqvist L, Hopkins B. Head position preference in the human newborn: a new look. Child Dev 1998 Feb;69(1):13-23. [doi: 10.1111/j.1467-8624.1998.tb06129.x] [Medline: 9499553]
49. Losee JE, Mason AC, Dudas J, Hua LB, Mooney MP. Nonsynostotic occipital plagiocephaly: factors impacting onset, treatment, and outcomes. Plast Reconstr Surg 2007 May;119(6):1866-1873. [doi: 10.1097/01.prs.0000259190.56177.ca] [Medline: 17440367]
50. Glasgow TS, Siddiqi F, Hoff C, Young PC. Deformational plagiocephaly: development of an objective measure and determination of its prevalence in primary care. J Craniofac Surg 2007 Jan;18(1):85-92. [doi: 10.1097/01.scs.0000244919.69264.bf] [Medline: 17251842]
51. van Vlimmeren LA, Helders PJ, van Adrichem LN, Engelbert RH. Torticollis and plagiocephaly in infancy: therapeutic strategies. Pediatr Rehabil 2006;9(1):40-46. [doi: 10.1080/13638490500037904] [Medline: 16352505]
52. Greve KR, Goldsbury CM, Simmons EA. Infants with congenital muscular torticollis requiring supplemental physical therapy interventions. Pediatr Phys Ther 2022 Jul 01;34(3):335-341. [doi: 10.1097/PEP.0000000000000906] [Medline: 35616484]
53. Ardic C, Usta O, Omar E, Yıldız C, Memis E. Effects of infant feeding practices and maternal characteristics on early childhood obesity. Arch Argent Pediatr 2019 Feb 01;117(1):26-33 [FREE Full text] [doi: 10.5546/aap.2019.eng.26] [Medline: 30652443]
54. Ross CE, Mirowsky J. The interaction of personal and parental education on health. Soc Sci Med 2011 Feb;72(4):591-599 [FREE Full text] [doi: 10.1016/j.socscimed.2010.11.028] [Medline: 21227556]
55. Moon RY, LoCasale-Crouch J, Turnbull KL, Colson E, Kellams A, Heeren T, et al. Investigating mechanisms for maternal education disparities in enacting health-promoting infant care practices. Acad Pediatr 2020;20(7):926-933. [doi: 10.1016/j.acap.2020.03.008] [Medline: 32201345]
56. Marchac A, Arnaud E, Di Rocco F, Michienzi J, Renier D. Severe deformational plagiocephaly: long-term results of surgical treatment. J Craniofac Surg 2011 Jan;22(1):24-29. [doi: 10.1097/SCS.0b013e3181f7dd4a] [Medline: 21187754]
57. Lennartsson F, Nordin P. Nonsynostotic plagiocephaly: a child health care intervention in Skaraborg, Sweden. BMC Pediatr 2019 Feb 06;19(1):48 [FREE Full text] [doi: 10.1186/s12887-019-1405-y] [Medline: 30727985]
58. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [Medline: 10877304]
59. Bialocerkowski AE, Vladusic SL, Wei Ng C. Prevalence, risk factors, and natural history of positional plagiocephaly: a systematic review. Dev Med Child Neurol 2008 Aug;50(8):577-586 [FREE Full text] [doi: 10.1111/j.1469-8749.2008.03029.x] [Medline: 18754894]

60. De Bock F, Braun V, Renz-Polster H. Deformational plagiocephaly in normal infants: a systematic review of causes and hypotheses. Arch Dis Child 2017 Jun;102(6):535-542. [doi: 10.1136/archdischild-2016-312018] [Medline: 28104626]
61. Inchingolo AA, Inchingolo AM, Piras F, Malcangi G, Patano A, Di Pede C, et al. A systematic review of positional plagiocephaly prevention methods for patients in development. Applied Sciences 2022 Nov 04;12(21):11172. [doi: 10.3390/app122111172]
62. Martiniuk AL, Vujovich-Dunn C, Park M, Yu W, Lucas BR. Plagiocephaly and developmental delay: a systematic review. J Dev Behav Pediatr 2017 Jan;38(1):67-78. [doi: 10.1097/DBP.0000000000000376] [Medline: 28009719]
63. Rohde JF, Goyal NK, Slovin SR, Hossain J, Pachter LM, Di Guglielmo MD. Association of positional plagiocephaly and developmental delay within a primary care network. J Dev Behav Pediatr 2021;42(2):128-134 [FREE Full text] [doi: 10.1097/DBP.0000000000000860] [Medline: 33538454]
64. Fontana SC, Daniels D, Greaves T, Nazir N, Searl J, Andrews BT. Assessment of deformational plagiocephaly severity and neonatal developmental delay. J Craniofac Surg 2016 Nov;27(8):1934-1936. [doi: 10.1097/SCS.0000000000003014] [Medline: 28005729]
65. Andrews BT, Fontana SC. Correlative vs. causative relationship between neonatal cranial head shape anomalies and early developmental delays. Front Neurosci 2017;11:708 [FREE Full text] [doi: 10.3389/fnins.2017.00708] [Medline: 29311788]
66. Ha EK, Lee SW, Kim JH, Lee E, Cha HR, Han BE, et al. Associations between delayed introduction of complementary foods and childhood health consequences in exclusively breastfed children. Nutrients 2023 Jul 31;15(15):3410 [FREE Full text] [doi: 10.3390/nu15153410] [Medline: 37571347]
67. Rasmussen SH, Shrestha S, Bjerregaard LG, Ängquist LH, Baker JL, Jess T, et al. Antibiotic exposure in early life and childhood overweight and obesity: a systematic review and meta-analysis. Diabetes Obes Metab 2018 Jun 25;20(6):1508-1514. [doi: 10.1111/dom.13230] [Medline: 29359849]
68. Ha EK, Kim JH, Park D, Lee E, Lee SW, Jee HM, et al. Personal exposure to total VOC is associated with symptoms of atopic dermatitis in schoolchildren. J Korean Med Sci 2022 Feb 28;37(8):e63 [FREE Full text] [doi: 10.3346/jkms.2022.37.e63] [Medline: 35226421]


## Abbreviations

**aOR:** adjusted odds ratio
**DDH:** developmental dysplasia of the hip
**DP:** deformational plagiocephaly
**GRADE:** Grading of Recommendations Assessment, Development, and Evaluation
**OR:** odds ratio
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RR:** risk ratio


*Edited by T de Azevedo Cardoso; submitted 20.12.23; peer-reviewed by SS Alemu, SW Lee, M Hikosaka; comments to author 14.05.24; revised version received 29.05.24; accepted 29.07.24; published 18.09.24.*

*Please cite as:*
*Hillyar CRT, Bishop N, Nibber A, Bell-Davies FJ, Ong J*
*Assessing the Evidence for Nonobstetric Risk Factors for Deformational Plagiocephaly: Systematic Review and Meta-Analysis*
*Interact J Med Res 2024;13:e55695*
*URL: https://www.i-jmr.org/2024/1/e55695*
*doi:10.2196/55695*
*PMID:39292504*

Review

# Health Locus of Control and Medical Behavioral Interventions: Systematic Review and Recommendations

Sogol Mozafari[1], MSIS; Alan Yang[1], PhD; Jason Talaei-Khoei[1], PhD

Department of Information Systems, College of Business, University of Nevada, Reno, Reno, NV, United States

**Corresponding Author:**
Alan Yang, PhD
Department of Information Systems
College of Business
University of Nevada, Reno
1664 North Virginia Street
Reno, NV, 89557
United States
Phone: 1 7757846428
Email: alany@unr.edu

## Abstract

**Background:** Health locus of control (HLOC) is a theory that describes how individuals perceive different forces that influence their lives. The concept of a locus of control can affect an individual's likelihood to commit to behaviors related to their health. This study explores the literature on the relationships between HLOC and medical behavioral interventions.

**Objective:** This study aims to better understand how HLOC constructs can potentially affect patient responses to health behavioral interventions and to propose a series of guidelines for individuals interested in designing medical behavioral interventions related to HLOC.

**Methods:** We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology and performed an analysis of 50 papers related to the topic of HLOC and medical behavioral interventions. Inclusion criteria were studies that had a behavioral intervention involving patients and contained a metric of at least 1 of the constructs related to HLOC. The initial screening and search were conducted by 2 researchers (AY and SM) separately. The results were then combined and compared.

**Results:** Our findings explore the influence of different levels of HLOC along with the importance of both patient- and health-related context when assessing the relationships between HLOC constructs and the likelihood of health behavior change. The findings show that different constructs related to HLOC can act as reliable predictors for patient responses to medical behavioral interventions. Patients who score higher on internal HLOC measures are more likely to exhibit behaviors that are consistent with positive health outcomes. Patients who score higher on chance HLOC are more likely to exhibit behaviors that may lead to adverse health outcomes. These conclusions are supported by most of the 50 studies surveyed.

**Conclusions:** We propose guidelines for individuals designing medical behavioral interventions so that they can make use of these relationships linked to HLOC. The three guidelines suggested are as follows: (1) in most situations, improving internal HLOC will improve health outcomes for patients; (2) patients with high external HLOC should be further studied to determine the source of the external HLOC; and (3) patients with a high chance HLOC are less likely to follow preventative behaviors or be responsive to interventions. Limitations of the study are that the primary search and analysis were conducted by 2 principal researchers (AY and SM). Interpretation and development of the guidelines are subject to individual interpretation of results and may not be applicable to all contexts.

***(Interact J Med Res 2024;13:e52287)*** doi:10.2196/52287

**KEYWORDS**

medical behavioral interventions; health locus of control; internal control; external control; patient behavior; synthesis; review methods; literature review; narrative review; behavior change

# Introduction

## Overview

Health locus of control (HLOC) refers to an individual's beliefs about the extent to which they have control over their health outcomes. Loci of control (LOCs) are typically split into 2 categories: internal and external forces. Individuals who have a higher internal LOC tend to believe that their actions can influence their health outcomes directly. Individuals with a higher external LOC are more likely to believe that their health outcomes are determined by external factors such as luck, fate, or the actions of other people. HLOC has been used as a theoretical model to predict the likelihood of individuals performing health-related behaviors across multiple settings. Medical behavioral interventions describe a broad range of activities; examples include the following: increasing the likelihood of individual vaccination rates through health messaging, increasing an individual's willingness to follow a meal plan and diet by communicating the benefits of weight loss, and increasing medication adherence through targeted messaging provided by a physician to a patient.

Medical behavioral interventions are meant to affect the health behaviors of individuals with the intention of improving health outcomes. HLOC is often used as a lens to interpret how an individual interprets and internalizes health-related stimuli. Health messaging, medication efficacy, and behavioral changes are all factors that have relationships with HLOC constructs [1,2]. Understanding an individual's HLOC can therefore be an important factor in determining the most effective approach to medical behavioral interventions.

## Research Question

The research question we are seeking to address is how do HLOC constructs affect the efficacy of health behavioral interventions? To explore this question, we conducted a systematic review of academic studies and randomized controlled trials related to health behavioral interventions and HLOC constructs. In addition to this research question, we also seek to develop a series of guidelines for individuals interested in designing medical behavioral interventions related to HLOC. The next section will provide background on the research space, specifically the HLOC and behavioral interventions in the health care context.

## Background

### *HLOC Theory*

The theory of LOC was developed in the 1950s by Julian Rotter [3]. Rotter separated the LOCs into 2 constructs: internal control and external control. The distinction between the 2 constructs is the extent to which an individual believes that the outcome of an event is "contingent on their own behavior" or "a function of chance, luck, or fate" [4,5]. HLOC is an extension of the LOC theory and relates to how much control people believe they have over their health. HLOC can also be either internal or external. HLOC extends the idea of external control by creating additional constructs to better define the source of external control. Studies on external HLOC (E HLOC) typically split the construct into 3 distinct sources: powerful others HLOC (PO HLOC), chance HLOC (C HLOC), and god HLOC (G HLOC) [2,6]. Internal HLOC (I HLOC) represents the belief of individuals in their ability to impact their health status, while E HLOC represents individual belief that external sources affect personal health [7].

Table 1 [4] describes the constructs of the LOC theory we focus on in this study. Internal and external LOCs are used to distinguish the individual constructs described in the table.

**Table 1.** Locus of control.

| Locus of control | Description |
|---|---|
| Internal | Individuals believe in their impact on their behavioral outcomes |
| External: powerful others | The belief of external sources and factors influencing one's life and decisions |
| External: chance | The belief of chance and fate influence one's life and decisions |
| External: god | The belief through religion of a higher power influencing one's life and decisions |

The HLOC constructs are independent of each other and can be applied to different contexts. People with higher I HLOC tend to believe in their influence on their behavioral outcomes, while individuals who have higher E HLOC believe that factors beyond their control affect those outcomes more strongly [8]. The impact of I HLOC can be effective toward physical self-care [9] and may also be an influential motivation relating to participation and the use of technologies or other tools designed for health care [5,10,11]. E HLOC and PO HLOC can be significant especially when it comes to the use of provider-recommended digital tools [1]. Individuals with I HLOC and PO HLOC beliefs may be more willing to use mobile health apps and platforms and are more likely to participate in monitoring health behaviors and developing web-based trackers [1]. HLOC has been used as a framework for analyzing individual health-related behaviors, particularly those related to platform use and trust.

HLOC remains a relevant theory for analyzing individual behavior within a health context. During the COVID-19 pandemic, many individuals experienced increases in factors related to I HLOC such as anxiety, fear, and depression. This increase led to many individuals experiencing posttraumatic stress disorder [8]. I HLOC is a commonly used framework for assessing individual factors influencing stress and developing solutions to manage anxiety [8]. E HLOC describes the extent to which an individual believes outside forces affect their life. The construct of powerful others in a health care context is

usually a measurement of individual trust in a health-related platform, app, or doctor [12].

### *Medical Behavioral Interventions*

Medical behavioral interventions refer to a wide range of strategies for the promotion of healthy behaviors in individual patients with the goal of improving health care outcomes [13-15]. Interventions can take on different forms including communicating health information to patients, providing incentives for desired behavior, or directly administering medication. Behavioral interventions tend to combine communication of information along with incentivization of healthy behaviors such as dieting, exercise, medication adherence, or cessation of substance abuse [16-19].

The efficacy of behavioral interventions depends on many factors. The desired health outcome and the demographics of the patient have large effects on the efficacy of different forms of interventions. For example, social support has been demonstrated to improve patient adherence to behavioral interventions related to lifestyle change such as dieting or exercise [20,21]. Interventions that are focused on acute diseases can change the efficacy of social incentives. Studies on behavioral interventions and cancer show that the severity of the disease affects adherence to guidelines delivered through an information-based health intervention [22-24]. Patients who are in the early stages of cancer tend to react strongly to new sources of information and seek out more when prompted, whereas patients in the latter stages are more receptive to the information provided by a trusted source such as a physician [25]. The consistency with which patients follow behavioral interventions is also not guaranteed. Most interventions tend to be administered over a period between 6 and 12 months. Past the intervention period, there is no certainty that patients who were receptive to the intervention messaging continued with the recommended behaviors [26-28].

In response to these issues, research has suggested that interventions tailored to specific individuals and their needs are more suitable than generalized approaches for behavioral change [29-32]. HLOC is a theory that is specific to individuals. An individual's perception of the world and what affects their health is an indicator of how they will respond to different forms of intervention [33-35]. Exploring this avenue of research can help uncover relationships between individual perceptions and the predictive efficacy of health interventions that would otherwise be unclear.

## Methods

### Overview

To accomplish our research objectives, we approached a systematic review of the literature using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology (Multimedia Appendix 1). PRISMA is a common methodology used to synthesize literature and filter sample pools in preparation for analysis [36-38]. Specifics regarding the search are reported below.

### PRISMA Search

We used the following terms when searching for papers related to our review: "Health Locus of Control"; "Behavior Change"; "Behavioral Change"; "Behavioral Change Interventions"; "Medication Adherence"; "HLOC"; "Patients"; "Healthcare; "Culture"; "Empirical"; and "Locus of Control." The terms were searched individually, then altogether using the keyword of "OR" connecting all of the different terms. Finally, the terms were searched together using the keyword "AND." These individual searches also included linked terms with "AND": "Health Locus of Control" AND "Behavior Change"; "Health Locus of Control" AND "Behavioral Change"; "Health Locus of Control" AND "Behavioral Change Interventions." The spelling of "Behaviour" was also used in the search for completeness. These keywords were used in a title plus abstract search through the databases of Web of Science, IEEE Xplore, the ACM Digital Library, the AIS e-Library, and PubMed. The search was conducted between November and December 2022. Additional details regarding the search and search strategy are available in Multimedia Appendix 2.

Inclusion criteria were studies that had a behavioral intervention involving patients and contained a metric of at least 1 of the constructs related to HLOC. Only studies that had actual data were considered; proposals and simulations were excluded. Additionally, only studies that were published in peer-reviewed journals were considered for inclusion. The initial screening and search were conducted by 2 researchers (AY and SM) separately. The results were then combined and compared. The initial search identified 357 papers potentially relevant to the study. Only references were used as a check for this stage, no abstracts or full texts of papers were read. After additional filtering, 245 redundant papers were removed from this pool. Reasons for paper removal during this stage included duplication due to cross-referencing and singular papers that appeared in multiple databases. Papers were then further checked for relevancy to the research scope. Filtering at this stage occurred from a reading of paper abstracts for relevance. If a study was not relevant based on a read of the abstract because it did not deal with medical behavioral intervention or HLOC or contain actual data, it was excluded. Papers that proposed studies but did not conduct behavioral interventions were also excluded at this stage. This resulted in the removal of an additional 72 papers, with 50 remaining for analysis. During this stage, citation software Zotero (Corporation for Digital Scholarship) was used to collect and compare citations and documentation.

### Data Extraction and Data Synthesis

Data extraction occurred after the identification of the items to be included in the analysis. Data extracted included the type of study, sample size, and interaction effects related to HLOC and the medical behavioral intervention. The full text of papers filtered through the first steps was read. Extraction was conducted by 2 independent researchers (AY and SM). Discrepancies were solved through a discussion of the 2 researchers and consensus was required to include information in the final document. Multimedia Appendix 3 [1,5,8,12,22,23,25,27,39-80] contains the results of this data extraction process. Data synthesis took on a similar form to

extraction, 2 independent researchers (AY and SM) evaluated the information collected and discrepancies were handled through consensus decision after discussion. After collating the results from the review, we analyzed the frequencies of studies that had relationships between the primary HLOC constructs and related health care behaviors. From this analysis, we developed a series of guidelines for researchers interested in the field of patient behavioral change in health care settings.

## Results

### Overview

The following sections will cover the results of the data synthesis. Topics of discussion include relationships between constructs, behavioral health interventions, and the influence of contextual factors such as disease and patient characteristics observed within the studies. Figure 1 contains the PRISMA flow diagram summarizing the search process.

The main health behaviors discussed in all the papers reviewed in this study are labeled into 5 main categories. These categories are preventative health behaviors, mental health, personal health perception, vaccination hesitancy, and physician trust. The number of papers that have focused on each health behavior is shown in Table 2.

The number of HLOC constructs that have been focused on in the 50 papers reviewed for this study can be seen in Table 3. Some studies have focused on several constructs, and some have focused on only one. The following table shows the counts of HLOC constructs that are mainly highlighted in this study's review papers.

**Figure 1.** PRISMA Search Strategy and Results.

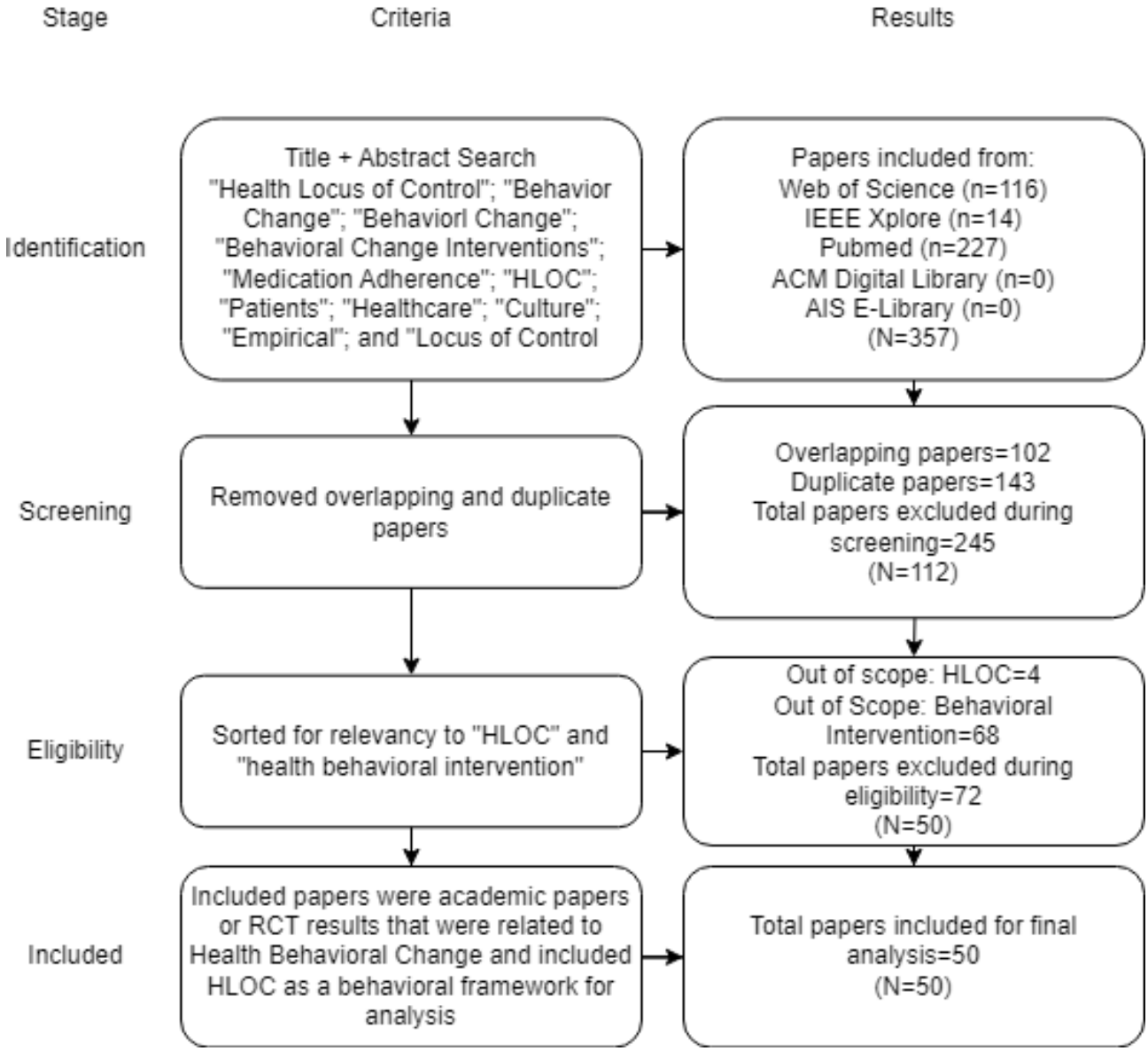


**Table 2.** Health behaviors.

| Health behavior | Papers (n=50), n (%) |
|---|---|
| Preventative health behaviors | 22 (44) |
| Mental health | 20 (40) |
| Personal health perception | 5 (10) |
| Vaccination hesitancy | 2 (4) |
| Physician trust | 1 (2) |

**Table 3.** Number of HLOC[a] constructs.

| HLOC constructs | Uses (n=50), n (%) |
|---|---|
| Internal HLOC | 45 (90) |
| Powerful others HLOC | 37 (74) |
| Chance HLOC | 34 (68) |
| God HLOC | 9 (18) |

[a]HLOC: health locus of control.

## HLOC and General Health

The relationship between HLOC and general patient health depends on many different factors. However, common relationships between HLOC constructs and patient health can be observed that are generalizable across contexts. The discussion in this section will focus on the I HLOC, E HLOC, PO HLOC, and C HLOC.

Higher I HLOC generally leads to better health outcomes overall due to higher rates of patient self-efficacy. Higher patient self-efficacy leads to an increase in behaviors that are beneficial to overall patient health. These behaviors can include improved medication adherence, preventative health behaviors, and an increase in healthy activities. These behaviors and outcomes are observable across income levels and other demographic characteristics [39,40].

Higher E HLOC within individuals can have different outcomes depending on the prevailing construct: C HLOC or PO HLOC. Higher C HLOC generally leads to more negative health behaviors and health outcomes in individuals. Individuals with high C HLOC tend to believe that health is an uncontrollable outcome that is determined by chance. As a result, those individuals tend to perform fewer preventative behaviors that could improve health outcomes. PO HLOC has a less clear relationship with general health. The source of the powerful other has a large impact on the decision an individual makes. For instance, individuals with a high PO HLOC who see physicians as the primary source of powerful others are likely to adhere to medical recommendations and have better health outcomes as a result. Individuals who have similarly high PO HLOC but view friends and family as the powerful other may have mixed results, as the friends and family may not recommend behaviors that are consistent with medical recommendations.

## HLOC and Health Interventions

Research has shown that higher I HLOC can lead to more preventative behavior from patients [41]. Patients with higher PO HLOC have more potential influences [21]. In fact, those with a higher PO HLOC are more likely to engage in preventative behaviors, particularly if the influence comes from physicians or other health care professionals [42]. However, if the PO HLOC comes from religion or a belief in a higher power, health care behaviors are likely more context dependent and less predictable [43]. For example, a high PO HLOC sourced to a patient's family may cause contradictory health outcomes such as lower vaccination rates, higher social contact, and higher physical activity. On the other hand, those with a higher C HLOC, or the belief that health outcomes are beyond their control, are less likely to engage in preventative health behaviors, often believing that their efforts won't matter.

Studies have found that individuals with higher I HLOC are more likely to engage in healthy dietary behaviors. This includes less consumption of unhealthy foods and more intake of healthy foods. On the other hand, those with higher C HLOC are less likely to adopt beneficial dietary patterns [44]. Interestingly, individuals with higher PO HLOC tend to have more beneficial dietary patterns. For example, parents can be seen as power others, and if an individual has a higher PO HLOC, they may be more likely to adopt a healthy diet based on their parents' influence [44]. Overall, an individual's HLOC plays a significant role in their dietary behaviors and can be a useful factor to consider when developing effective interventions for improving nutrition.

When it comes to substance abuse, research suggests that HLOC can play a significant role in this context. Individuals with higher I HLOC tend to make self-centered decisions that can lead to reduced substance abuse. This includes decisions such as smoking reduction, birth control use, and weight loss [9]. Additionally, those with lower I HLOC are less likely to make these self-centered decisions and tend to act more carelessly toward their health [45]. Examples would be more alcohol

consumption and daily smoking. Similarly, individuals with a higher C HLOC have been found to be more likely to smoke. Conversely, those with a higher PO HLOC are less likely to smoke due to the influence of their family or physicians [45]. Understanding an individual's HLOC can be an important factor in developing effective interventions to reduce substance abuse and improve overall health outcomes.

### Additional Factors

Besides the individual contexts of the relationships between constructs, individual patient characteristics can also affect how HLOC constructs impact health behaviors. Patients who have chronic conditions tend to have higher C HLOC, particularly over longer periods. The severity of illnesses also affects the usual relationships between HLOC constructs and patient health. In the context of chronic conditions, such as asthma, patients with higher self-efficacy and higher I HLOC are in more control of their condition, whereas patients with higher E HLOC beliefs can potentially lead to poorer control [46].

### Patient Characteristics

Several factors have been found to influence HLOC, including social contact, health information–seeking behavior, and trust in health care providers. For example, social contact has been found to lead to better health care behaviors across all income levels [39]; higher I HLOC has been found to lead to more social behaviors, which in turn leads to positive health influence [47]; and health information–seeking behavior has been linked to higher I HLOC, indicating individuals with higher I HLOC are more likely to seek out information to be well aware of their health decisions outcome [39]. Higher E HLOC, and mostly higher PO HLOC, can increase patients' level of trust, and patients who are more trusting have been found to be more likely to follow their physician recommendations [12]. C HLOC has also been found to have a moderating effect on patient self-efficacy and self-management behavior [48]. High self-efficacy leads to more self-management, and depending on the level of patients' C HLOC, this relationship can be either strengthened or weakened.

E HLOC has been found to affect death-related anxiety. Higher PO HLOC can lead to less stress regarding the uncontrollable nature of death based on the trust in others' (physicians') influence on one's health outcomes [49]. However, higher C HLOC has been found to increase the level of anxiety and depression at the end of life, as it indicates patients' lack of control over their health and the inevitability of death [22]. These patterns are also applicable to healthy patients as they age (Aviad and Cohen-Louck [72]). E HLOC can also impact patients both positively and negatively. An elevated E HLOC can result in an increased level of depression and anxiety for a patient. For example, patients with high E HLOC have been found to have poor coping skills, which could result in higher levels of depression. Within the same context, patients with high E HLOC linked to caregivers are more likely to rely on their physician, worry less, and adapt to their conditions [22,81].

### Caregiver Characteristics

The impact of an individual's health decisions on other individuals and their relationship with them can also affect their perceived HLOC. For example, if a patient is concerned about their child's health outcome, the parent's I HLOC and E HLOC may be influenced by their perception of their ability and connected anxiety related to their child's health outcomes. For example, parental caregiving quality increases when a child faces challenges such as a chronic condition [27]. Some studies have found mothers to be the main caregivers in special chronic conditions such as type 1 diabetes [82,83]. Parents of children dealing with diabetes were found to have higher I HLOC, whereas patients who had failed to reach that control level tended to have lower levels of I HLOC and higher levels of G HLOC and C HLOC [82]. Mothers and other caregivers have also been found to influence patients' behavior and coping styles. The level of I HLOC or E HLOC of the caregiver can significantly impact the patients' active or passive coping style and how they behave toward their condition and perceived anxiety level [84].

### Disease Characteristics

Terminal illnesses can lead to unique cases where higher I HLOC causes greater stress and anxiety for patients as they experience a greater loss of control [22,50]. The relationship between these 2 constructs is stronger as the illness becomes more serious. A reversal of this relationship is possible. Patients who are provided with an option for treatment or hopeful messaging can decrease stress and anxiety among patients with higher I HLOC. The likely reason for this phenomenon is that patients with higher I HLOC perceive a return of control, which leads to lower stress and more proactive behaviors when positive messaging is provided.

Patients with cancer across all stages tend to have higher PO HLOC and decreased I HLOC. Higher PO HLOC tends to manifest in a lowering of anxiety if the source of the powerful other is rooted in religion or part of the G HLOC [25]. The reverse of this relationship has also been observed. Patients with cancer with higher I HLOC tend to have lower E HLOC. The higher I HLOC is usually indicative of a higher quality of life for the duration of the disease [51]. This might be due to the impact of I HLOC in the hope of recovery, which leads the patient to a better mental state and lowers the anxiety and depression rates.

Uncertainty is a major consideration across multiple studies focused on patient behaviors. Uncertainty is defined as any situation where a decision maker is unable to accurately predict outcomes because they lack information or sufficient environmental cues [85]. In the concept of HLOC, individuals with higher I HLOC are more likely to react to uncertainty as a controllable decision and opportunity, whereas individuals with higher E HLOC may consider their uncertainty as untrustworthy evidence of a situation [86]. Patients with a higher level of I HLOC tend to be more involved in their health-related situations and have less uncertainty about them [51,87,88]. Higher I HLOC and more level of uncertainty in patients can result in less stress and anxiety rates.

Research shows that uncertainty mediates between HLOC with quality of life, anxiety, and depression [51]. It has been illustrated that the level of a patient's education and the complexity of a treatment can directly affect patients'

uncertainty which can be reduced by improving the level of education and patients' knowledge about the related phenomenon [89-91]. Physicians and health professionals can be external resources for providing such information to patients with higher PO HLOC and help them overcome their uncertainty and improve their mental health [91,92]. Patients with higher I HLOC have been found to have better coping with their condition and adapt well to the change, along with information-seeking behavior, which would result in improved level of knowledge about their condition and reduced level of stress [51,87]. Additionally, patients who have dealt with chronic conditions for longer periods feel less uncertainty and less stress about their condition because of their perceived knowledge about their situation over time which would result in their improved quality of life and mental health. With chronic conditions, such as asthma, patients with higher self-efficacy and higher I HLOC are in more control of their condition, whereas patients with higher E HLOC beliefs can potentially lead to poorer control [46]. In the case of patients with dementia, higher E HLOC has been observed with lower patient depression, which could be because E HLOC may be linked to acceptance of chronic conditions (Halse et al [50]).

# Discussion

## Principal Findings

Through analysis of the data, we observed intersections between HLOC characteristics of patients and health outcomes. The most commonly recurrent themes are related to I HLOC, E HLOC, and C HLOC. The effects of these constructs varied across studies. Based on common patterns observed across the data, we offer the following suggestions for individuals designing medical behavioral interventions. First, in most situations, improving I HLOC will improve health outcomes for patients. I HLOC has been observed to improve patient health outcomes across many different contexts. Patients with higher I HLOC feel a greater sense of ownership over their health, which leads them to take more proactive measures and makes them more receptive to health behavioral interventions. The one exception to this guideline is situations where a patient's health condition is terminal and an individual has limited control over the outcome. In such situations, high I HLOC can be problematic and lead to higher stress in contexts where patients feel a loss of control [88]. In these situations, cultivating E HLOC or G HLOC or minimizing I HLOC through a discussion around acceptance of outcomes and a release of personal control may be beneficial [51]. Second, patients with high E HLOC should be further studied to determine the source of the E HLOC. A common misconception of patients with high E HLOC is that they are not as receptive to health messaging because they are not intrinsically motivated like individuals with high I HLOC. Individuals with high E HLOC can be just as receptive and proactive as individuals with high I HLOC, but the source of the external influence is an important factor to consider. Patients who perceive their physicians or similar caregivers as the source of the E HLOC will be more receptive to following suggestions provided by the perceived authority [3]. However, patients who identify their main E HLOC source as friends, family, or media may hold beliefs that are contradictory to positive health messaging [52]. Determining the source of E HLOC is crucial to a better understanding of how those patients are likely to respond to health interventions. Third, patients with high C HLOC are less likely to follow prepreventative behaviors or be responsive to interventions. Across all studies, contexts, and situations surveyed for this review, patients with higher C HLOC were less prone to be proactive about their health. This likely stems from the idea common to individuals with demonstrable high C HLOC that actions are ultimately meaningless because everything happens through chance including health outcomes. To address these issues, health interventionalists should identify patients who exhibit high C HLOC and seek to educate them on the benefits of behavioral change. Demonstrations on the effectiveness of interventions can also help to decrease C HLOC and improve healthy patient behaviors.

## Comparison to Prior Work

We explored the relationship between HLOC and health behavioral interventions in this paper. Previous studies have explored similar relationships related to individual-level constructs and responses to health behavioral interventions and health messaging. In this section, we explore some of the prior work in the literature and how it relates to individual-level beliefs regarding personal health.

HLOC and more generally LOC as a whole pertain to a high degree of conceptual similarities with the attributional notion of locus of cause, which refers to individuals' perception of their causation [93]. The concepts, however, differ in the way a specific behavior is observed: either from an outside observer's perspective or one observing his or her behavior [94]. The locus of cause is considered internal when someone can perform a task and tries to do so. In this case, failing the task is attributed to a lack of trying and effort of the person. In contrast, when someone does not have the required abilities for a specific task, the locus of cause appears to be external, and the related failure is assumed to be due to external circumstances [93,94]. Similarly, the theory of self-attribution bias highlights a situation in which individuals excessively attribute credit to their abilities for past successes, while assigning blame to others for failures [95-97]. Based on this concept, individuals attribute themselves to positive outcomes, linking them to their actions and efforts, while ascribing negative results to external factors such as bad luck [98]. In the context of LOC, individuals either perceive themselves to have the ability to control an outcome and try to do so or perceive themselves as not having the ability to control an outcome and therefore do not put any effort into achieving that. It can be concluded that the LOC can be simplified as a way of attributing the cause and outcomes of one's actions to oneself [94].

Studies have posited that older people and those with lower levels of education have higher levels of E HLOC beliefs over internal ones [40]. In addition to education, various studies have reported higher E HLOC in individuals with low sociodemographic status or negative health-related behavior such as smoking and drinking [1,99-104]. This association has been specifically investigated by Wallston and Wallston [105], which highlighted higher PO HLOC and higher C HLOC in individuals with less than 12 years of education. Conversely,

higher I HLOC can be seen in individuals with high socioeconomic status or positive health behavior such as regular exercising or staying on a healthy diet [53]. On the same stream of thoughts, the HLOC theory has been considered as a possible mediator between socioeconomic status and health outcomes [53,106].

### Limitations and Strengths of the Systematic Review

The systematic review has inherent limitations. Observation of a field of research is dependent upon the scope and perception of the researchers cataloging the information. Analysis and interpretation of results are affected by subjective factors. In this paper, we have attempted to minimize individual biases and subjectivity by establishing inclusion and exclusion criteria and having sourcing and analysis of data conducted by 2 researchers (AY and SM) independently. The strengths of the systematic review are the ability to provide a comprehensive view of the state of research and the ability to generate guidelines that are related to current trends and recent phenomena in the field.

### Future Work

This study focused on the literature surrounding HLOC constructs and health behavioral interventions. New health behavioral interventions related to the COVID-19 pandemic and the subsequent policies offer new avenues for the exploration of HLOC and individual health behaviors. Future studies could explore how social distancing, vaccinations, or quarantining is related to the idea of HLOC and the propensity of individuals to adhere to guidelines. In general, it is probable for individuals with higher PO HLOC to be influenced by other individuals' suggestions. However, individuals with higher I HLOC are also likely to get this influence from others depending on the situation. A relative example of this context is the nonforcing manipulation through nudges. Nudging entails organizing and modifying a decision-making scenario without restricting choice options or imposing force, in a way to alter someone's behavior in a predictive way [107]. For instance, a smoker with high I HLOC may decide to quit smoking due to a close friend telling him or her about the infertility risks caused by smoking. In this example, the individual's decision about willingly stopping an internal choice with the intention of maintaining his or her health is taking place because of someone else's nudge. In other words, this person believes in having self-control over this decision and is aware of having the choice of not quitting as well but is still influenced to do so. Many examples can be applicable in this context, depending on the specific scenario and individuals' circumstances. We recommend future research to expand on the impact of "compelling versus noncompelling manipulation," as well as "effective nudges from friends or family versus strangers" on HLOC.

An exploration of the individual sources of E HLOC can also be beneficial. E HLOC is often interpreted in a similar manner to I HLOC. However, the sources of E HLOC are myriad and can vary from individual to individual. More studies that explore how changes to different HLOC constructs affect patient behaviors over time would support the relationships observed in many of the papers. One additional avenue of research is the implementation of health interventions with artificial intelligence. Research on this topic has already explored the possibility of integrating language models or similar generative agents with adaptive responses to improve patient health outcomes [108,109]. The guidelines suggested in this paper could also be tested in a longitudinal study where changes in HLOC construct strength could be measured against patient behaviors in response to behavioral interventions.

This study's methodology is a comprehensive literature review. Studies were filtered and analyzed based on relevance but there was no metric applied to rate the quality of the studies or to determine the suitability of the findings beyond the existence of data collection and analysis. Future studies could apply more stringent metrics toward inclusion of studies, such as by omitting studies that do not exceed a certain sample size or only including studies that meet a particular duration.

## Conclusions

The study of HLOC and medical behavioral interventions is the study of how each individual responds to messaging and motivations regarding their health-related behaviors. An increased understanding of the constructs and relationships across these 2 ideas can lead to better-designed studies and interventions across different populations. Further exploration of this topic can focus on the importance of individual characteristics and the influence of context on each HLOC construct. The patterns and relationships frequently observed can help both academics and practitioners better design studies to explore questions related to improving health outcomes. Better-designed interventions can lead to individuals taking a more active role in managing their health, ultimately leading to improved health outcomes for everyone.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist for systematic literature review.
[PDF File (Adobe PDF File), 304 KB - ijmr_v13i1e52287_app1.pdf ]

Multimedia Appendix 2
Search strategy for systematic literature review.

[DOCX File , 15 KB - ijmr_v13i1e52287_app2.docx ]

Multimedia Appendix 3
Notes for systematic literature review.
[DOCX File , 68 KB - ijmr_v13i1e52287_app3.docx ]

## References


1. Bennett BL, Goldstein CM, Gathright EC, Hughes JW, Latner JD. Internal health locus of control predicts willingness to track health behaviors online and with smartphone applications. Psychol Health Med 2017;22(10):1224-1229 [FREE Full text] [doi: 10.1080/13548506.2017.1317354] [Medline: 28415852]
2. Kassianos AP, Symeou M, Ioannou M. The health locus of control concept: factorial structure, psychometric properties and form equivalence of the multidimensional health locus of control scales. Health Psychol Open 2016;3(2):2055102916676211 [FREE Full text] [doi: 10.1177/2055102916676211] [Medline: 35223073]
3. Rotter J. Rotter internal-external locus of control scale. 28 Meas Locus Control 2011;10:1-97 [FREE Full text]
4. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr Gen Appl 1966;80(1):1-28. [doi: 10.1037/h0092976]
5. Ahadzadeh AS, Wu SL, Ong FS, Deng R. The mediating influence of the unified theory of acceptance and use of technology on the relationship between internal health locus of control and mobile health adoption: cross-sectional study. J Med Internet Res 2021;23(12):e28086 [FREE Full text] [doi: 10.2196/28086] [Medline: 34964718]
6. Wallston KA, Wallston BS, DeVellis R. Development of the multidimensional health locus of control (MHLC) scales. Health Educ Monogr 1978;6(2):160-170. [doi: 10.1177/109019817800600107] [Medline: 689890]
7. Wallston KA. The validity of the multidimensional health locus of control scales. J Health Psychol 2005;10(5):623-631. [doi: 10.1177/1359105305055304] [Medline: 16033784]
8. Chandra Y, Yagnik J. Experience of perceived stress and impact of health locus of control during COVID-19 pandemic: investigating entrepreneurs and corporate employees. South Asian J Hum Resour Manag 2022;9(1):79-99. [doi: 10.1177/23220937221076245]
9. Wallston BD, Wallston KA. Locus of control and health: a review of the literature. Health Educ Monogr 1978;6(2):107-117. [doi: 10.1177/109019817800600102] [Medline: 357347]
10. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
11. Yang A, Varshney U. Mobile health evaluation: taxonomy development and cluster analysis. Healthc Anal 2022;2:100022. [doi: 10.1016/j.health.2022.100022]
12. Hillen MA, de Haes HCJM, Stalpers LJA, Klinkenbijl JHG, Eddes EH, Verdam MGE, et al. How attachment style and locus of control influence patients' trust in their oncologist. J Psychosom Res 2014;76(3):221-226. [doi: 10.1016/j.jpsychores.2013.11.014] [Medline: 24529041]
13. Lippke S, Ziegelmann JP. Theory-based health behavior change: developing, testing, and applying theories for evidence-based interventions. Appl Psychol 2008;57(4):698-716. [doi: 10.1111/j.1464-0597.2008.00339.x]
14. Ryan P. Integrated theory of health behavior change: background and intervention development. Clin Nurse Spec 2009;23(3):161-170; quiz 171-172 [FREE Full text] [doi: 10.1097/NUR.0b013e3181a42373] [Medline: 19395894]
15. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
16. Lu Y, Stathopoulou T, Vasiloglou MF, Pinault LF, Kiley C, Spanakis EK, et al. goFOODTM: an artificial intelligence system for dietary assessment. Sensors (Basel) 2020;20(15):4283 [FREE Full text] [doi: 10.3390/s20154283] [Medline: 32752007]
17. MacPherson MM, Merry KJ, Locke SR, Jung ME. Effects of mobile health prompts on self-monitoring and exercise behaviors following a diabetes prevention program: secondary analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2019;7(9):e12956 [FREE Full text] [doi: 10.2196/12956] [Medline: 31489842]
18. Nabovati E, Vakili-Arki H, Taherzadeh Z, Saberi MR, Medlock S, Abu-Hanna A, et al. Information technology-based interventions to improve drug-drug interaction outcomes: a systematic review on features and effects. J Med Syst 2017;41(1):12. [doi: 10.1007/s10916-016-0649-4] [Medline: 27889873]
19. Borghouts J, Eikey E, Mark G, de Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
20. Koufopoulos JT, Conner MT, Gardner PH, Kellar I. A web-based and mobile health social support intervention to promote adherence to inhaled asthma medications: randomized controlled trial. J Med Internet Res 2016;18(6):e122 [FREE Full text] [doi: 10.2196/jmir.4963] [Medline: 27298211]
21. Marr J, Wilcox S. Self-efficacy and social support mediate the relationship between internal health locus of control and health behaviors in college students. Am J Health Educ 2015;46(3):122-131. [doi: 10.1080/19325037.2015.1023477]

22. Brown AJ, Thaker PH, Sun CC, Urbauer DL, Bruera E, Bodurka DC, et al. Nothing left to chance? The impact of locus of control on physical and mental quality of life in terminal cancer patients. Support Care Cancer 2017;25(6):1985-1991. [doi: 10.1007/s00520-017-3605-z] [Medline: 28175997]
23. Lima MP, Moret-Tatay C, Irigaray TQ. Locus of control, personality and depression symptoms in cancer: testing a moderated mediation model. Clin Psychol Psychother 2022;29(2):489-500. [doi: 10.1002/cpp.2604] [Medline: 33908672]
24. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
25. Iskandarsyah A, de Klerk C, Suardi DR, Sadarjoen SS, Passchier J. Health locus of control in Indonesian women with breast cancer: a comparison with healthy women. Asian Pac J Cancer Prev 2014;15(21):9191-9197 [FREE Full text] [doi: 10.7314/apjcp.2014.15.21.9191] [Medline: 25422200]
26. Kang Y. Role of health locus of control between uncertainty and uncertainty appraisal among patients with atrial fibrillation. West J Nurs Res 2009;31(2):187-200. [doi: 10.1177/0193945908326063] [Medline: 19050229]
27. Tsiouli E, Pavlopoulos V, Alexopoulos EC, Chrousos G, Darviri C. Short-term impact of a stress management and health promotion program on perceived stress, parental stress, health locus of control, and cortisol levels in parents of children and adolescents with diabetes type 1: a pilot randomized controlled trial. Explore (NY) 2014;10(2):88-98. [doi: 10.1016/j.explore.2013.12.004] [Medline: 24607075]
28. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. JMIR Hum Factors 2015;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
29. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first health information national trends survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
30. Hill TD, Upenieks L, Ellison CG. Religious involvement, health locus of control, and sleep disturbance: a study of older Mexican Americans. In: Angel JL, Ortega ML, Robledo LMG, editors. Understanding the Context of Cognitive Aging: Mexico and the United States. Cham, Switzerland: Springer International Publishing; 2021:161-179.
31. Xie H, Prybutok G, Peng X, Prybutok V. Determinants of trust in health information technology: an empirical investigation in the context of an online clinic appointment system. Int J Human-Computer Interact 2020;36(12):1095-1109. [doi: 10.1080/10447318.2020.1712061]
32. Perez-Ramos JG, McIntosh S, Barrett ES, Velez Vega CM, Dye TD. Attitudes toward the environment and use of information and communication technologies to address environmental health risks in marginalized communities: prospective cohort study. J Med Internet Res 2021;23(9):e24671 [FREE Full text] [doi: 10.2196/24671] [Medline: 34554103]
33. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
34. Campbell RJ, Nolfi DA. Teaching elderly adults to use the internet to access health care information: before-after study. J Med Internet Res 2005;7(2):e19 [FREE Full text] [doi: 10.2196/jmir.7.2.e19] [Medline: 15998610]
35. Howell P, Abdelhamid M. Protection motivation perspective regarding the use of COVID-19 mobile tracing apps among public users: empirical study. JMIR Form Res 2023;7:e36608 [FREE Full text] [doi: 10.2196/36608] [Medline: 36735838]
36. Pinho S, Cruz M, Ferreira F, Ramalho A, Sampaio R. Improving medication adherence in hypertensive patients: a scoping review. Prev Med 2021;146:106467. [doi: 10.1016/j.ypmed.2021.106467] [Medline: 33636195]
37. Singh N, Varshney U. IT-based reminders for medication adherence: systematic review, taxonomy, framework and research directions. Eur J Inf Syst 2019;29(1):84-108. [doi: 10.1080/0960085x.2019.1701956]
38. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
39. Musich S, Wang SS, Slindee L, Kraemer S, Yeh CS. The impact of internal locus of control on healthcare utilization, expenditures, and health status across older adult income levels. Geriatr Nurs 2020;41(3):274-281 [FREE Full text] [doi: 10.1016/j.gerinurse.2019.10.008] [Medline: 31727348]
40. Poortinga W, Dunstan FD, Fone DL. Health locus of control beliefs and socio-economic differences in self-rated health. Prev Med 2008;46(4):374-380. [doi: 10.1016/j.ypmed.2007.11.015] [Medline: 18177930]
41. Bianchi D, Lonigro A, Norcia AD, Tata DD, Pompili S, Zammuto M, et al. A model to understand COVID-19 preventive behaviors in young adults: health locus of control and pandemic-related fear. J Health Psychol 2022;27(14):3148-3163 [FREE Full text] [doi: 10.1177/13591053221089722] [Medline: 35410516]
42. Marton G, Pizzoli SFM, Vergani L, Mazzocco K, Monzani D, Bailo L, et al. Patients' health locus of control and preferences about the role that they want to play in the medical decision-making process. Psychol Health Med 2021;26(2):260-266 [FREE Full text] [doi: 10.1080/13548506.2020.1748211] [Medline: 32323553]
43. Olagoke AA, Olagoke OO, Hughes AM. Intention to vaccinate against the novel 2019 coronavirus disease: the role of health locus of control and religiosity. J Relig Health 2021;60(1):65-80 [FREE Full text] [doi: 10.1007/s10943-020-01090-9] [Medline: 33125543]

44. Duplaga M, Grysztar M. Nutritional behaviors, health literacy, and health locus of control of secondary schoolers in Southern Poland: a cross-sectional study. Nutrients 2021;13(12):4323 [FREE Full text] [doi: 10.3390/nu13124323] [Medline: 34959875]
45. Mercer DA, Ditto B, Lavoie KL, Campbell T, Arsenault A, Bacon SL. Health locus of control is associated with physical activity and other health behaviors in cardiac patients. J Cardiopulm Rehabil Prev 2018;38(6):394-399. [doi: 10.1097/HCR.0000000000000350] [Medline: 30106790]
46. Schreitmüller J, Loerbroks A. The role of self-efficacy and locus of control in asthma-related needs and outcomes: a cross-sectional study. J Asthma 2020;57(2):196-204. [doi: 10.1080/02770903.2018.1556687] [Medline: 30634885]
47. Kesavayuth D, Poyago-Theotoky J, Tran DB, Zikos V. Locus of control, health and healthcare utilization. Econ Model 2020;86:227-238. [doi: 10.1016/j.econmod.2019.06.014]
48. Wang R, Zhou C, Wu Y, Sun M, Yang L, Ye X, et al. Patient empowerment and self-management behaviour of chronic disease patients: a moderated mediation model of self-efficacy and health locus of control. J Adv Nurs 2022;78(4):1055-1065. [doi: 10.1111/jan.15077] [Medline: 34643959]
49. Akanni O, Koleoso O. Self-esteem, locus of control and religiosity in predicting death anxiety among students in a Nigerian tertiary institution. J Clin Sci Res 2022;11(1):28-33. [doi: 10.4103/jcsr.jcsr_35_21]
50. Halse I, Bjørkløf GH, Engedal K, Selbæk G, Barca ML. One-year change in locus of control among people with dementia. Dement Geriatr Cogn Dis Extra 2021;11(3):298-305 [FREE Full text] [doi: 10.1159/000520248] [Medline: 35082825]
51. Sharif SP. Locus of control, quality of life, anxiety, and depression among Malaysian breast cancer patients: the mediating role of uncertainty. Eur J Oncol Nurs 2017;27:28-35. [doi: 10.1016/j.ejon.2017.01.005] [Medline: 28279393]
52. Boyd JM, Wilcox S. Examining the relationship between health locus of control and god locus of health control: is god an internal or external source? J Health Psychol 2020;25(7):931-940. [doi: 10.1177/1359105317739099] [Medline: 29115175]
53. Kuwahara A, Nishino Y, Ohkubo T, Tsuji I, Hisamichi S, Hosokawa T. Reliability and validity of the multidimensional health locus of control scale in Japan: relationship with demographic factors and health-related behavior. Tohoku J Exp Med 2004;203(1):37-45 [FREE Full text] [doi: 10.1620/tjem.203.37] [Medline: 15185970]
54. Ganjoo M, Farhadi A, Baghbani R, Daneshi S, Nemati R. Association between health locus of control and perceived stress in college student during the COVID-19 outbreak: a cross-sectional study in Iran. BMC Psychiatry 2021;21(1):529 [FREE Full text] [doi: 10.1186/s12888-021-03543-1] [Medline: 34702220]
55. Lee DJ, So WY, Lee SM. The relationship between Korean adolescents' sports participation, internal health locus of control, and wellness during COVID-19. Int J Environ Res Public Health 2021;18(6):2950 [FREE Full text] [doi: 10.3390/ijerph18062950] [Medline: 33805732]
56. Shin S, Lee E. Relationships among the internal health locus of control, mental health problems, and subjective well-being of adults in South Korea. Healthcare (Basel) 2021;9(11):1588 [FREE Full text] [doi: 10.3390/healthcare9111588] [Medline: 34828633]
57. Imeri H, Holmes E, Desselle S, Rosenthal M, Barnard M. A survey study of adults with chronic conditions: examining the correlation between patient activation and health locus of control. Chronic Illness 2023;19(1):118-131. [doi: 10.1177/17423953211067431] [Medline: 36638782]
58. Mehta R, Narayanan M. The relationship of emotional reactivity with health locus of control. J Psychosocial Res 2021;162(1):47-57 [FREE Full text]
59. Aviad-Wilchek Y. Locus of control and the meaning of life as a salutogenic model that reduces suicidal tendencies in patients with mental illness. Curr Psychol 2019;40(2):465-474. [doi: 10.1007/s12144-018-0122-2]
60. Wrightson KJ, Wardle J. Cultural variation in health locus of control. Ethn Health 1997;2(1-2):13-20. [doi: 10.1080/13557858.1997.9961811] [Medline: 9395585]
61. Reknes G, Visockaite A, Liefooghe A, Lovakov A, Einarsen SV. Locus of control moderates the relationship between exposure to bullying behaviors and psychological strain. Front Psychol 2019;10:1323 [FREE Full text] [doi: 10.3389/fpsyg.2019.01323] [Medline: 31244725]
62. Mahmoud AB, Reisel WD, Fuxman L, Hack‐Polay D. Locus of control as a moderator of the effects of COVID‐19 perceptions on job insecurity, psychosocial, organisational, and job outcomes for MENA region hospitality employees. Eur Manag Rev 2021;19(2):313-332. [doi: 10.1111/emre.12494]
63. Helmer SM, Krämer A, Mikolajczyk RT. Health-related locus of control and health behaviour among university students in North Rhine Westphalia, Germany. BMC Res Notes 2012;5:703 [FREE Full text] [doi: 10.1186/1756-0500-5-703] [Medline: 23273039]
64. Amit Aharon A, Nehama H, Rishpon S, Baron-Epel O. A path analysis model suggesting the association between health locus of control and compliance with childhood vaccinations. Hum Vaccin Immunother 2018;14(7):1618-1625 [FREE Full text] [doi: 10.1080/21645515.2018.1471305] [Medline: 29771633]
65. Morishita M, Hattori S, Miyai N. Ability for self-care among elderly patients with diabetes mellitus and its association with health locus of control and social support. Nihon Eiseigaku Zasshi 2017;72(1):77-86 [FREE Full text] [doi: 10.1265/jjh.72.77] [Medline: 28154365]
66. Kaynak H, Turan A, Demir Y. Locus of control as a mediator of the relationships between motivational systems and trait anxiety. Psychol Rep 2024;127(4):1533-1560. [doi: 10.1177/00332941221139707] [Medline: 36377649]

67. Krampe H, Danbolt LJ, Haver A, Stålsett G, Schnell T. Locus of control moderates the association of COVID-19 stress and general mental distress: results of a Norwegian and a German-speaking cross-sectional survey. BMC Psychiatry 2021;21(1):437 [FREE Full text] [doi: 10.1186/s12888-021-03418-5] [Medline: 34488667]
68. Micheletto V, Zito M, Bustreo M, Gabrielli G, Circi R, Russo V. The impact of optimism and internal locus of control on workers' well-being, a multi-group model analysis before and during the COVID-19 pandemic. Social Sci 2022;11(12):559. [doi: 10.3390/socsci11120559]
69. Mori M, Seko T, Ogawa S. Association of social capital and locus of control with perceived health during the COVID-19 pandemic in Japan. Int J Environ Res Public Health 2022;19(15):9415 [FREE Full text] [doi: 10.3390/ijerph19159415] [Medline: 35954772]
70. Weinhardt CB, Ruckert JH. Internal locus of control predicts proenvironmental and COVID-19 health-related behaviors: a pilot study. Ecopsychology 2023;15(4):368-381. [doi: 10.1089/eco.2022.0016]
71. Würtzen H, Clausen LH, Andersen PB, Santini ZI, Erkmen J, Pedersen HF. Mental well-being, health, and locus of control in Danish adults before and during COVID-19. Acta Neuropsychiatr 2022;34(2):93-98. [doi: 10.1017/neu.2021.37] [Medline: 34814958]
72. Aviad Y, Cohen-Louck K. Locus of control and purpose in life as protective factors against the risk for suicide in older adults. Smith Coll Stud Social Work 2021;91(4):295-308. [doi: 10.1080/00377317.2021.1968323]
73. Gore JS, Griffin DP, McNierney D. Does internal or external locus of control have a stronger link to mental and physical health? Psychol Stud 2016;61(3):181-196. [doi: 10.1007/s12646-016-0361-y]
74. Grotz M, Hapke U, Lampert T, Baumeister H. Health locus of control and health behaviour: results from a nationally representative survey. Psychol Health Med 2011;16(2):129-140. [doi: 10.1080/13548506.2010.521570] [Medline: 21328142]
75. Moshki M, Amiri M, Khosravan S. Mental health promotion of Iranian university students: the effect of self-esteem and health locus of control. J Psychiatr Ment Health Nurs 2012;19(8):715-721. [doi: 10.1111/j.1365-2850.2011.01806.x] [Medline: 22074491]
76. Holroyd EA, Molassiotis A, Taylor-Pilliae RE. Filipino domestic workers in Hong Kong: health related behaviors, health locus of control and social support. Women Health 2001;33(1-2):181-205. [doi: 10.1300/J013v33n01_11] [Medline: 11523638]
77. Cohen M, Azaiza F. Health-promoting behaviors and health locus of control from a multicultural perspective. Ethn Dis 2007;17(4):636-642. [Medline: 18072372]
78. Pieterse AL, Carter RT. An exploratory investigation of the relationship between racism, racial identity, perceptions of health, and health locus of control among Black American women. J Health Care Poor Underserved 2010;21(1):334-348. [doi: 10.1353/hpu.0.0244] [Medline: 20173273]
79. Egan JT, Leonardson G, Best LG, Welty T, Calhoun D, Beals J. Multidimensional health locus of control in American Indians: the strong heart study. Ethn Dis 2009;19(3):338-344 [FREE Full text] [Medline: 19769018]
80. van Dijk TK, Dijkshoorn H, van Dijk A, Cremer S, Agyemang C. Multidimensional health locus of control and depressive symptoms in the multi-ethnic population of the Netherlands. Soc Psychiatry Psychiatr Epidemiol 2013 Dec;48(12):1931-1939. [doi: 10.1007/s00127-013-0678-y] [Medline: 23508370]
81. Broers S, Kaptein AA, Le Cessie S, Fibbe W, Hengeveld MW. Psychological functioning and quality of life following bone marrow transplantation: a 3-year follow-up study. J Psychosom Res 2000;48(1):11-21. [doi: 10.1016/s0022-3999(99)00059-8] [Medline: 10750625]
82. Franceschi R, Canale M, Piras EM, Galvagni L, Vivori C, Cauvin V, et al. Influence of parental health locus of control on behavior, self-management and metabolic control, in pediatric patients with type 1 diabetes. J Pers Med 2022;12(10):1590 [FREE Full text] [doi: 10.3390/jpm12101590] [Medline: 36294729]
83. Malerbi FEK, Negrato CA, Gomes MB, Brazilian Type 1 Diabetes Study Group (BrazDiab1SG). Assessment of psychosocial variables by parents of youth with type 1 diabetes mellitus. Diabetol Metab Syndr 2012;4(1):48 [FREE Full text] [doi: 10.1186/1758-5996-4-48] [Medline: 23174044]
84. Polizzi C, Perricone G, Fontana V, D'Angelo P, Jankovic M, Nichelli F, et al. The relation between maternal locus of control and coping styles of pediatric leukemia patients during treatment. Pediatr Rep 2020;12(1):7998 [FREE Full text] [doi: 10.4081/pr.2020.7998] [Medline: 32308968]
85. Mishel MH, Clayton MF. Theories of uncertainty in illness. In: Middle Range Theory for Nursing. 3rd Edition. New York, NY: Springer Publishing Company; 2008:53-86.
86. Mishel MH. Uncertainty in illness. Image J Nurs Sch 1988;20(4):225-232. [doi: 10.1111/j.1547-5069.1988.tb00082.x] [Medline: 3203947]
87. Lin CC, Tsay HF. Relationships among perceived diagnostic disclosure, health locus of control, and levels of hope in Taiwanese cancer patients. Psychooncology 2005;14(5):376-385. [doi: 10.1002/pon.854] [Medline: 15386760]
88. Watson M, Pruyn J, Greer S, van den Borne B. Locus of control and adjustment to cancer. Psychol Rep 1990;66(1):39-48. [doi: 10.2466/pr0.1990.66.1.39] [Medline: 2326428]
89. Ko NY, Hsu ST. Informational needs, health locus of control and uncertainty among women hospitalized with gynecological diseases. Chang Gung Med J 2005;28(8):559-566 [FREE Full text] [Medline: 16265846]

90. Mishel MH. Reconceptualization of the uncertainty in illness theory. Image J Nurs Sch 1990;22(4):256-262. [doi: 10.1111/j.1547-5069.1990.tb00225.x] [Medline: 2292449]
91. Mishel MH. Uncertainty in acute illness. Annu Rev Nurs Res 1997;15:57-80. [Medline: 9262787]
92. van der Molen B. Relating information needs to the cancer experience: 1. Information as a key coping strategy. Eur J Cancer Care (Engl) 1999;8(4):238-244. [doi: 10.1046/j.1365-2354.1999.00176.x] [Medline: 10889622]
93. Heider F. The psychology of interpersonal relations. Psychology Press. 1958. URL: https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780203781159&type=googlepdf [accessed 2024-01-02]
94. Stewart JE. Locus of control and self-attribution as mediators of hazardous attitudes among aviators: a review and suggested applications. Int J Appl Aviat Stud 2008;85(2):263-279 [FREE Full text]
95. Campbell WK, Sedikides C. Self-threat magnifies the self-serving bias: a meta-analytic integration. Rev Gen Psychol 1999;3(1):23-43. [doi: 10.1037//1089-2680.3.1.23]
96. Miller DT, Ross M. Self-serving biases in the attribution of causality: fact or fiction? Psychol Bull 1975;82(2):213-225. [doi: 10.1037/h0076486]
97. Zuckerman M. Attribution of success and failure revisited, or: the motivational bias is alive and well in attribution theory. J Pers 1979;47(2):245-287. [doi: 10.1111/j.1467-6494.1979.tb00202.x]
98. Stewart JE. Locus of control, attribution theory, and the "Five Deadly Sins" of aviation. U.S. Army Research Institute for the Behavioral and Social Sciences 2006:1-35. [doi: 10.1037/e584572011-001]
99. Calnan M. Control over health and patterns of health-related behaviour. Soc Sci Med 1989;29(2):131-136. [doi: 10.1016/0277-9536(89)90161-5] [Medline: 2749298]
100. Fleming MF, Barry KL. Health locus of control in a primary care sample of alcoholics and nonalcoholics. Behav Med 1991;17(1):25-30. [doi: 10.1080/08964289.1991.9937549]
101. Paxton SJ, Sculthorpe A. Weight and health locus of control beliefs in an Australian community sample. Psychol Health 2007;14(3):417-431. [doi: 10.1080/08870449908407338]
102. Schank MJ, Lawrence DM. Young adult women: lifestyle and health locus of control. J Adv Nurs 1993;18(8):1235-1241. [doi: 10.1046/j.1365-2648.1993.18081235.x] [Medline: 8376662]
103. Slenker SE, Price JH, O'Connell JK. Health locus of control of joggers and nonexercisers. Percept Mot Skills 1985;61(1):323-328. [doi: 10.2466/pms.1985.61.1.323] [Medline: 4047893]
104. Stuart K, Borland R, McMurray N. Self-efficacy, health locus of control, and smoking cessation. Addict Behav 1994;19(1):1-12. [doi: 10.1016/0306-4603(94)90046-9] [Medline: 8197887]
105. Wallston KA, Wallston BS. Health locus of control scales. Res Locus Control Constr 1981;1(1):189. [doi: 10.1016/b978-0-12-443201-7.50010-5]
106. Acosta MSV, Chapman P, Bigelow PL, Kennedy C, Buchan RM. Measuring success in a pesticide risk reduction program among migrant farmworkers in Colorado. Am J Ind Med 2005;47(3):237-245. [doi: 10.1002/ajim.20136] [Medline: 15712260]
107. Hausman DM, Welch B. Debate: to nudge or not to nudge. J Polit Philos 2010;18(1):123-136 [FREE Full text]
108. Kim J, Lee D, Park E. Machine learning for mental health in social media: bibliometric study. J Med Internet Res 2021;23(3):e24870 [FREE Full text] [doi: 10.2196/24870] [Medline: 33683209]
109. Asan O, Choudhury A. Research trends in artificial intelligence applications in human factors health care: mapping review. JMIR Hum Factors 2021;8(2):e28236 [FREE Full text] [doi: 10.2196/28236] [Medline: 34142968]


## Abbreviations

**C HLOC:** chance health locus of control
**E HLOC:** external health locus of control
**G HLOC:** god health locus of control
**HLOC:** health locus of control
**I HLOC:** internal health locus of control
**LOC:** locus of control
**PO HLOC:** power others health locus of control
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

*Edited by T de Azevedo Cardoso; submitted 29.08.23; peer-reviewed by S Subramaniam, T Martin, A Palanica; comments to author 19.01.24; revised version received 14.02.24; accepted 26.06.24; published 10.10.24.*

*Please cite as:*
*Mozafari S, Yang A, Talaei-Khoei J*
*Health Locus of Control and Medical Behavioral Interventions: Systematic Review and Recommendations*
*Interact J Med Res 2024;13:e52287*
*URL: https://www.i-jmr.org/2024/1/e52287*
*doi:10.2196/52287*
*PMID:*

Review

# Blended Psychological Therapy for the Treatment of Psychological Disorders in Adult Patients: Systematic Review and Meta-Analysis

Kelly Ferrao Nunes-Zlotkowski[1,2], BPsych (Hon), MSGC; Heather L Shepherd[1,3], BA (Hons), Dip HE, PhD; Lisa Beatty[1,4], Bpsych (Hons), PhD; Phyllis Butow[1,2], BA (Hons), DipEd, MClin Psych, MPH, PhD; Joanne Margaret Shaw[1,2], BAppl Sc, BPsych (Hon), PhD

[1]Psycho-Oncology Co-operative Research Group (PoCoG), School of Psychology, The University of Sydney, Sydney, Australia

[2]Centre for Medical Psychology & Evidence-based Decision-making (CeMPED), School of Psychology, The University of Sydney, Sydney, Australia

[3]Susan Wakil School of Nursing and Midwifery, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[4]Flinders University, College of Education, Psychology and Social Work, Flinders University Institute of Mental Health and Wellbeing, Adelaide, Australia

**Corresponding Author:**
Joanne Margaret Shaw, BAppl Sc, BPsych (Hon), PhD
Psycho-Oncology Co-operative Research Group (PoCoG)
School of Psychology
The University of Sydney
Rm 311B Level 3 Griffith Taylor (A19)
Manning Road
Sydney, NSW 2006
Australia
Phone: 61 2 9351 3761
Email: joanne.shaw@sydney.edu.au

## Abstract

**Background:** Blended therapy (BT) combines digital with face-to-face psychological interventions. BT may improve access to treatment, therapy uptake, and adherence. However, research is scarce on the structure of BT models.

**Objective:** We synthesized the literature to describe BT models used for the treatment of psychological disorders in adults. We investigated whether BT structure, content, and ratio affected treatment efficacy, uptake, and adherence. We also conducted meta-analyses to examine treatment efficacy in intervention-control dyads and associations between treatment outcomes versus BT model structure.

**Methods:** PsycINFO, CINAHL, Embase, ProQuest, and MEDLINE databases were searched. Eligibility criteria included articles published in English till March 2023 that described digital and face-to-face elements as part of an intervention plan for treating psychological disorders in adult patients. We developed a coding framework to characterize the BT interventions. A meta-analysis was conducted to calculate effect size (ES; Cohen *d* and 95% CIs) regarding pre- and posttreatment outcomes in depression and anxiety versus BT structure. The review was registered with PROSPERO and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

**Results:** Searches identified 8436 articles, and data were extracted from 29 studies. BT interventions were analyzed and classified according to mode of interaction between digital and face-to-face components (*integrated* vs *sequential*), role of the components (*core* vs *supplementary*), component delivery (*alternate* vs *case-by-case*), and digital materials assignment mode (*standardized* vs *personalized*). Most BT interventions (n=24) used a cognitive behavioral therapy approach for anxiety or depression treatment. Mean rates of uptake (91%) and adherence (81%) were reported across individual studies. BT interventions were more effective or noninferior to treatment as usual, with large spread in the data and a moderate to large ES in the treatment of depression (n=9; Cohen *d*=–1.1, 95% CI –0.6 to –1.6, *P*<.001, and *z* score=–4.3). A small, nonsignificant ES was found for anxiety outcomes (n=5; Cohen *d*=–0.1, 95% CI –0.3 to 0.05, *P*=.17, and *z* score=–1.4). Higher ESs were found in blended interventions with supplementary design (depression: n=11, Cohen *d*=–0.75, 95% CI –0.56 to –0.95; anxiety: n=8, Cohen *d*=–0.9, 95% CI –0.6 to –1.2); fewer (≤6) face-to-face sessions (depression: n=9, Cohen *d*=–0.7, 95% CI –0.5 to –0.9; anxiety: n=7, Cohen *d*=–0.8, 95% CI –0.3 to –1.3); and a lower ratio (≤50%) of face-to-face versus digital sessions (depression: n=5, Cohen *d*=–0.8, 95% CI –0.6 to –1.1; anxiety: n=4, Cohen *d*=–0.8, 95% CI 0.006 to –1.6).

**Conclusions:** This study confirmed integrated BT models as feasible to deliver. We found BT to be effective in depression treatment, but anxiety treatment results were nonsignificant. Future studies assessing outcomes across different psychological disorders and therapeutic approaches are required.

**Trial Registration:** PROSPERO CRD42021258977; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=258977

***(Interact J Med Res 2024;13:e49660)*** doi:10.2196/49660

**KEYWORDS**

systematic review; blended psychological therapy; blended care; face-to-face; online; psychological intervention; intervention design; digital care; digital mental health; psychological disorder

## Introduction

### The Impact of Psychological Disorders

Psychological disorders affect approximately 1 billion people globally and were responsible for 10% of the prepandemic global burden of disease in 2019, with estimates of the mental health burden increasing [1]. Depression and anxiety disorders cost US $1 trillion per year globally [2]. Evidence-based, effective mental health care is available, but its provision does not reach everyone who needs it [2,3].

### Digital (or Online) Therapy

Digital psychological therapy provides an opportunity to enhance patient access to psychological treatment [4,5] and is recommended by the World Health Organization as a cornerstone of "comprehensive, integrated and responsive mental health and social care services" [1]. Research regarding digital therapy has largely focused on internet-delivered cognitive behavioral therapy (CBT), demonstrating its efficacy and cost-effectiveness [6,7], particularly in the treatment of depression and anxiety [8-12]. Although uptake and adherence to digital therapy in the research setting have shown improvement [13], engagement is still low, particularly in routine mental health care [14-17]. In addition, research suggests stakeholder preference for face-to-face interventions [18,19]. A systematic review [20] reported on concerns raised by health professionals about the use of digital therapy alone, including perceived nonsuitability for patients due to symptom severity, lack of digital access and literacy, and perception of digital treatment as being less engaging than face-to-face treatments. This review indicated that blended psychological therapy (also called "blended therapy" or "blended care") was perceived as a midway option between digital and face-to-face therapy.

### Blended Therapy

Blended therapy (BT) is a model of care that combines digital and face-to-face delivery of psychological therapy, integrating benefits from both modalities. Specifically, the face-to-face component is delivered by a mental health professional, such as a psychologist, while the digital component is patient driven [21-23]. Integrating digital therapy with face-to-face interventions in a blended model has the potential to save professionals' and patients' time (eg, transport to and from the clinic); increase the frequency of sessions; improve treatment uptake, adherence, and maintenance; and boost therapy effects [24-26].

A systematic review by Erbe et al [24] (N=44) found that BT may improve dropout rates and save health professionals' time compared with exclusively face-to-face interventions. Despite increasing evidence of the benefits of blended psychological therapy for patients [22,24], there is a lack of research specifically focused on "what, how, where, and when" BT is effective to inform future BT interventions [21,22,27]. The rationale for our systematic review emerges from the scarcity of data specifically focusing on BT processes including BT content and structure, which hinders scientific reproducibility of BT and impacts its implementation success.

### Objectives of This Review

Seeking to address these gaps in BT literature, our systematic review and meta-analysis expands on the work of Erbe et al [24] and aims to (1) identify and describe the structure, content, and ratio of the face-to-face and digital components in BT interventions applied for the treatment of psychological disorders and (2) investigate whether there is an association among the structure, content, and ratio of blended components and the treatment efficacy of, uptake of, and adherence to BT.

## Methods

### Design

This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [28] guidelines and was registered with PROSPERO (CRD42021258977). The PRISMA checklist is provided in Multimedia Appendix 1.

### Search Strategy

The PsycINFO, CINAHL, Embase, ProQuest, and MEDLINE databases were searched using keywords and Medical Subject Headings terms—"blended"; "online"; "face-to-face"; "treatment"; "therapy"; "care"; "mental disorders"; "psychological distress"; and "psychological disease"—for articles published in English (Multimedia Appendix 2). The search included articles published till May 2022, and an updated search was conducted in March 2023. Reference lists of the included studies were also manually searched.

### Study Selection

#### *Inclusion and Exclusion Criteria*

Studies that described or applied an intervention where both digital and face-to-face elements were integrated or delivered sequentially were included. We included studies in which the participants were aged ≥18 years and diagnosed with a psychological disorder. Studies solely investigating populations

other than this target group (health care professionals, student cohorts, employees, etc) were excluded.

### *Comparators*

The comparator or control groups included treatment as usual (pharmacological or psychological intervention and standard medical care), waitlist, or other interventions.

### *Data Abstracted*

The primary focus was the intervention design, including descriptions of the structure, content, and ratio of the sessions used in each model. Secondary outcomes were (1) a psychological therapy approach used in the BT models, (2) patient groups for which BT was applied, (3) treatment efficacy, (4) uptake and adherence, (5) health service outcomes (eg, cost-effectiveness), (6) patients' acceptability of BT, (7) therapeutic alliance rates, and (8) barriers and facilitators reported.

### *Article Screening and Selection*

All search results were uploaded into Covidence software [29]. Two reviewers (KFN-Z and JMS) screened the titles and abstracts independently. Full-text reviews based on the eligibility criteria were conducted by KFN-Z and PB or JMS.

## Data Extraction

Data extraction was conducted by KFN-Z using a purpose-designed data extraction template. Extraction results were partially reviewed (6/29, 20%) by a second coder (JMS) to assess accuracy. To capture any missing data, the corresponding authors were e-mailed twice. Data extracted included the following: (1) study characteristics—authors, year of publication, country, study setting, study aims, study type, sample size, control group (where applicable), therapy approach applied, primary or secondary psychological outcomes, and symptom assessment measures and participant characteristics such as age, sex, diagnosed psychological disorders, severity of symptoms, and individual study outcomes; (2) intervention characteristics—BT intervention design based on the structure, content, and ratio of BT sessions; number, periodicity, and duration (in minutes) of face-to-face and digital sessions; treatment length (in weeks); and (3) BT intervention outcomes—treatment efficacy, uptake, adherence, cost-effectiveness, acceptability, therapeutic alliance, and barriers and facilitators to BT reported.

## Data Analysis

To address the objectives of this review, quantitative variables regarding the BT intervention structure were summarized and described. We used descriptive statistics (mean, percentages, and range) to describe quantitative data regarding study and participant characteristics; BT intervention uptake, adherence, and completion rates; treatment length (in weeks); number, time (in minutes), ratio, and periodicity of face-to-face and digital sessions; treatment acceptability; efficacy; and therapeutic alliance. Barriers to and facilitators of BT were qualitatively analyzed using a thematic analysis [30] approach. Qualitative data on BT structure and content were analyzed using a content analysis approach [31]. Categories and subcategories were summarized in a framework that builds on the concepts described by Erbe et al [24] (Textbox 1).

**Textbox 1.** Classification—blended model designs.

**Interaction between face-to-face and digital components: *integrated* vs *sequential***

- *Integrated* models present both the digital and face-to-face components as collaborating parts within a therapy regimen, with both components delivered within the course of the intervention [24].
- *Sequential* models present the digital component delivered in entirety before or after face-to-face component delivery [24]. Sequential interventions start by delivering a "batch" of either face-to-face or digital sessions. Once the first "batch" is finished, the other component gets delivered.
    - *Stepped care* is considered a special type of *sequential* design in which the digital component is a step in the intervention sequence [24]. *Stepped care* interventions deliver the least intensive or costly treatment first and then progress to more intensive or aftercare treatment, if required. Hence, the "blend" in stepped care only effectuates after the first stage (digital) of treatment is complete and if patients require additional (face-to-face) care.

**Role of the components in the intervention: *core* vs *supplementary***

- *Core* components are an indispensable part of the blended intervention, as they deliver new therapeutic elements (ie, modules complement each other).
- *Supplementary* components present content that has already been discussed during the intervention, that is, content of one component is supplementary to the content delivered in the main component. For example, face-to-face content may be supplemented by reinforcing exercises and homework on the web.

**Delivery pattern of face-to-face or digital components: *alternate* vs *case by case***

- *Alternate* delivery is a configuration in which each session is delivered by alternating face-to-face or digital components in a fixed ratio. The distribution of components is preset for the entire intervention; this may feature as a 1:1 ratio, but other ratios (eg, 2 digital to1 face-to-face) of distribution are possible.
    - *Linear* delivery is specific to *sequential* designs in which all digital sessions are delivered in a row followed by all face-to-face sessions in a row, or vice versa.
- *Case-by-case* delivery is a configuration in which therapists assess and formulate a strategy for distributing the face-to-face or digital sessions adapted to the clients' or patients' needs on a case-by-case basis.

**Digital content assignment: *personalized* vs *standardized***

- *Personalized* content assignment is not preset; therapists and patients decide which materials to complete, tailoring them to patients' needs.
- *Standardized* content assignment is largely preset; materials are delivered to all patients undergoing treatment with little or no changes to content.

### *Meta-Analyses*

Although a meta-analysis was not originally included in the registered protocol, data collection and analysis processes indicated the relevance of meta-analyzing treatment outcomes for enhancing systematic review results. The meta-analysis was conducted using the Comprehensive Meta-Analysis program [32] to investigate treatment efficacy between the treatment and control dyads. A meta-analysis of *BT interventions only* was also conducted to investigate associations between BT structure and content and treatment outcomes. Pre-post outcome means and SD, alongside data on sample size at post–time points were included in the analysis. Standardized difference in means (Cohen *d*) and 95% CIs were used as effect size (ES) measures, and *z* values were used to test the null hypothesis (ES=0). We used a random effects meta-analysis due to expected heterogeneity. ES was set to negative numbers to show the change in symptoms (the lower the number, the higher the reduction in symptoms). Heterogeneity was assessed using *Q* test (*Q*), I-squared ($I^2$), Tau-squared ($T^2$), and Tau (T) scores. Publication bias was assessed using funnel plots and the trim and fill method by Duval and Tweedie [33].

### *Risk of Bias and Quality Assessment*

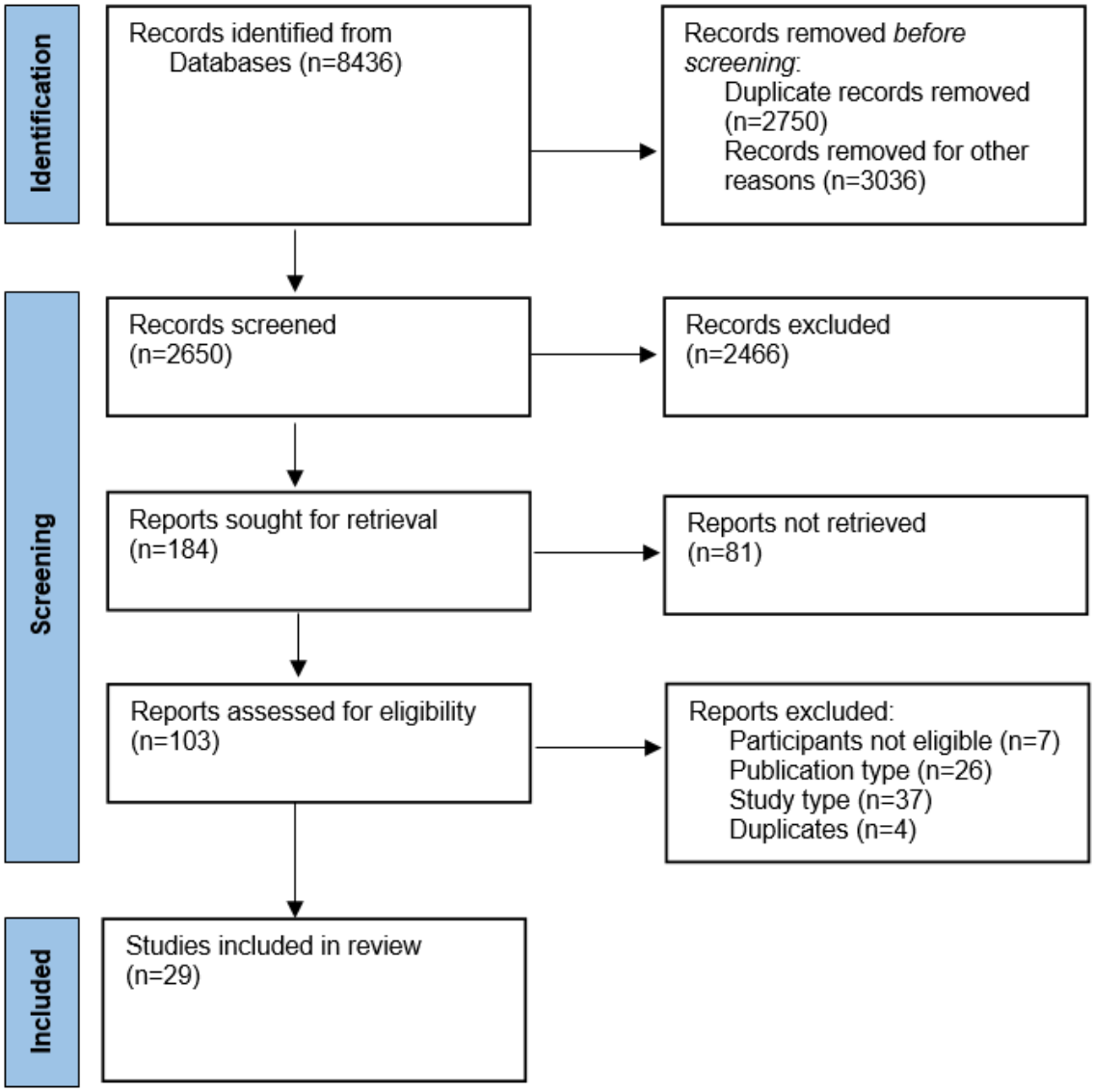
Two independent reviewers (KFN-Z and JMS) assessed the risk of bias using the National Institutes of Health Study Quality Assessment Tool (quantitative) [34], the Critical Appraisal Skills Programme qualitative checklist [35], or the Mixed Methods Appraisal Tool [36].

## Results

### Database Search Results

Database searches identified 2650 papers after removal of duplicates. Title and abstract screening resulted in 103 articles for full-text review. A total of 30 eligible articles were identified but only 29 were included in the review—one eligible paper was excluded as it reported the same data. The PRISMA diagram in Figure 1 [28] provides the details of the process.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram.




## Quality Assessment

Quality assessment deemed most studies (n=22) to be of "good quality" as they addressed most of the quality assessment criteria applied. Seven articles [23,37-42] were classified to be of "fair" quality (Multimedia Appendix 3).

## Study Characteristics

Of the 29 articles, 25 (86%) were prospective studies (randomized controlled trials [RCTs], n=14; feasibility or pilot, n=4; cohort or single arm, n=7); 3 (10%) were retrospective analyses of cohort studies; and 1 (3%) study used qualitative methods only. Most BT interventions (22/29, 76%) primarily treated depression, either exclusively (13/29, 45%) or in combination with anxiety treatment (9/29, 31%). Most studies (28/29, 96%) used a CBT approach, with 3 (10%) combining CBT and other approaches such as dialectical behavioral therapy and acceptance and commitment therapy (n=2) and motivational interviewing (n=1). Outcomes assessed were primarily symptom reduction (25/29, 86%), although 4 (14%) studies reported on the intervention process or working alliance outcomes as the primary focus (Table 1).

Collectively, the studies included a total of 12,322 (range 3-4448) patient participants, with 57.71% (n=7111) receiving BT interventions. Of the studies reporting on symptom reduction (28/29, 96%), all prescreened for clinical levels of psychological morbidity (Table 2).

**Table 1.** Characteristics of reviewed studies.

| Study; country | Study setting | Study aims | Study design | Participants, N; BT[a] group, n; comparator, n | Therapy type; clinical outcomes |
|---|---|---|---|---|---|
| Askjer and Mathiasen [43], 2021 and Mathiasen et al [44],[b] 2022; Denmark | Specialized mental health care | Explore if WA[c] predicted treatment outcomes | Exploratory; secondary analysis from an RCT[d] | 76; 38; 38; comparator: F2F[e] CBT[f] | CBT; WA, depressive symptoms |
| Berger et al [45], 2018; Germany | Routine outpatient psychotherapy practices | Investigate web-based treatment as adjunctive to depression treatment vs regular psychotherapy | Two-armed, pragmatic RCT | 98; 51; 47; comparator: TAU[g]: psychotherapy | CBT; depressive symptoms |
| Bisson et al [46], 2022; United Kingdom | Primary and secondary mental health settings | Determine if CBT-TF[h] was noninferior to F2F CBT-TF for PTSD[i] | Pragmatic, multicenter, noninferiority RCT | 196; 97; 99; comparator: F2F CBT-TF | CBT; severity of symptoms of PTSD |
| Cloitre et al [47], 2022; United States | Routine care in mental health outpatient clinics | Assess 2 ratios of F2F sessions to self-guided work on trauma-exposed veterans | Quasi-experiment with a noninferiority design | 202; 202; NC[j] | Transdiagnostic, trauma-informed CBT; PTSD and depression |
| Duffy et al [48], 2020; England | Specialized mental health care service | Investigate iCBT[k] as a prequel for high-intensity depression and anxiety treatment | Uncontrolled feasibility design (open study) | 123; 123; NC | CBT; anxiety and depression and work and social functioning |
| Etzelmueller et al [49], 2018; Germany | Routine care practice (clinics) | Evaluate patient experience of a blended iCBT service | Qualitative, semistructured interviews | 15; 15; NC | CBT; ND[l] depression |
| Høifødt et al [37], 2013; Norway | University outpatient clinic | Evaluate effectiveness and acceptability of a guided web-based program for depression | RCT | 106; 52; 54; comparator: delayed treatment, waitlist, or TAU | CBT; depression symptoms |
| Jacmon et al [42], 2009; Australia | Psychology private practice | Assess cost-effectiveness and convenience of partially digital depression treatment | Single-arm, pretest-posttest study | 9; 9; NC | CBT; depression levels |
| Kemmeren et al [50], 2019; France, Germany, Poland, and Netherlands | Multisetting (specialized mental health and routine primary care) | Examine use of and engagement to blended CBT for depression | Exploratory, secondary study from RCT | 231; 231; NC | CBT; ND depression |
| Kenter et al [51], 2013; Netherlands | Mental health care center; routine care | Report on the uptake of digital treatment, on the profile of patients who prefer digital therapy, and on symptom reduction vs waitlist | Observational study (electronic patient database) | 104; 55; 49; comparator: waitlist | PST[m]; ND depression, anxiety, and burnout |
| Kenter et al [52], 2015; Netherlands | Mental health service | Compare the effects and costs between blended and F2F treatments | Naturalistic study: examined records of patients | 4448; 168; 4280; comparator: TAU: F2F | CBT; ND depression and anxiety |
| Kok et al [38], 2014; Netherlands | Outpatient clinics | Assess clinical effectiveness of internet-based guided self-help vs waitlist | RCT | 212; 105; 107; comparator: waitlist | Psychotherapy; phobia and avoidance behavior |
| Kooistra et al [23], 2016; Netherlands | Outpatient specialized mental health care center | Develop and evaluate a structured, blended CBT protocol for patients with depression | Focus-groups and single-arm, pre-post | 30; 9; NC (12 patients) | CBT; ND depression |

| Study; country | Study setting | Study aims | Study design | Participants, N; BT[a] group, n; comparator, n | Therapy type; clinical outcomes |
|---|---|---|---|---|---|
| Kooistra et al [26], 2019; Netherlands | Specialized mental health care (outpatient services) | Compare costs and effectiveness of blended vs standard CBT for depression | Pilot RCT with 2 parallel groups | 102; 53; 49; comparator: CBT-F2F | CBT; self-reported depression severity |
| Kooistra et al [53], 2020; Netherlands | Outpatient specialized mental health clinics | Investigate WA in bCBT[n] for depression | Exploratory, secondary study from pilot RCT | 92; 47; 45; comparator: regular CBT | CBT; ND depression levels |
| Lungu et al [54], 2020; United States | Employer program | Evaluate the effectiveness of a video-based CBT and internet intervention | Retrospective cohort study | 385; 385; NC | CBT+UTP[o], ACT[p], and DBT[q]; ND depression and anxiety |
| Ly et al [55], 2015; Sweden | Clinical setting | Evaluate a blended treatment for depression | Noninferiority RCT | 93; 46; 47; comparator: F2F BA[r] | BA; depression |
| Månsson et al [39], 2013; Sweden | Clinical setting | Explore clinical outcomes and user experiences of internet-delivered therapy. To develop or test a bCBT model | Mixed methods, case series, pilot study (pre-post re BT testing) | 23; 15; NC (patients, n=15; therapists, n=8) | CBT; ND anxiety and depression |
| Månsson et al [40], 2017; Sweden | Outpatient psychiatric clinic | Evaluate an internet-based support as adjunct to F2F CBT | Feasibility study | 54; 45; NC (patients, n=45; therapists, n=9) | CBT; ND anxiety and depression symptoms |
| Mol et al [56], 2018; Netherlands | Outpatient clinic | Explore therapist behaviors; adherence; and patient outcome in digital therapy | Observational study | 64; 45; NC (patients, n=45; therapists, n=19) | CBT; ND depression levels |
| Nakao et al [57], 2018; Japan | Outpatient medical institutions | Evaluate effectiveness of web-based bCBT in reducing therapist time in patients with depression | Single-blinded RCT | 40; 20; 20; comparator: waitlist + pharmacological treatment | CBT; depression symptoms |
| Romijn et al [41], 2021 and Romijn et al [58];[b] Netherlands | Outpatient specialized mental health care centers | Explore therapist fidelity to bCBT protocols for anxiety disorders | Mixed methods (derived from a larger RCT) | 114; 52; 62; comparator: CBT F2F | CBT; anxiety symptoms |
| Tarp et al [59], 2022; Denmark | Public municipal outpatient alcohol clinics | Describe development and testing of a digital program; participant experiences; and usability of BT | Feasibility and pilot study | 32; 22; NC (development: 7 therapists+3 patients) | CBT + motivational interviewing; —[s] (intervention system usability) |
| Thase et al [60], 2018; United States | Department of psychiatry of medical schools | Evaluate the efficacy of computer and therapist-assisted CBT vs standard CBT | Noninferiority RCT | 154; 77; 77; comparator: CBT F2F | CBT; depression symptom severity |
| van de Wal et al [61], 2017; Netherlands | Cancer hospitals: academic, regional, and outpatient | Investigate the efficacy of BT for FCR[t] in cancer survivors | RCT: 2-arm, parallel group, longitudinal | 88; 45; 43; comparator: TAU (any) | CBT; FCR severity |
| Vernmark et al [62], 2019; Sweden | Mental health care centers | Explore patient- and therapist-rated WA in bCBT and WA as a predictor for change | Exploratory secondary study from RCTs | 151; 75; NC | CBT; depression levels |
| Witlox et al [63], 2021; Netherlands | Mental health service at general practices | Examine the effectiveness of blended ACT for older adults with anxiety | RCT (single-blinded) | 314; 150; 164; comparator: TAU (FTF CBT) | ACT; anxiety severity |

| Study; country | Study setting | Study aims | Study design | Participants, N; BT[a] group, n; comparator, n | Therapy type; clinical outcomes |
|---|---|---|---|---|---|
| Wu et al [64], 2021; United States | Employer mental health program clinical services | Evaluate the outcomes of a blended care coaching program for anxiety and depression | Retrospective cohort analysis | 1496; 1496; NC | CBT-based + CBT, DBT, and ACT; anxiety and depression symptoms |
| Wu et al [65], 2021; United States | Employer mental health program clinical services | Examine the effectiveness and the impact of bCBT on anxiety and depression | Retrospective cohort analysis | 3401; 3401; NC | CBT+DBT and ACT; anxiety and depression symptoms |

[a]BT: blended therapy.
[b]Linked study.
[c]WA: working alliance.
[d]RCT: randomized controlled trial.
[e]F2F: face to face.
[f]CBT: cognitive behavioral therapy.
[g]TAU: treatment as usual.
[h]iCBT-TF: internet-guided cognitive behavioral therapy with trauma focus.
[i]PTSD: posttraumatic stress disorder.
[j]NC: no comparator.
[k]iCBT: internet-delivered cognitive behavioral therapy.
[l]ND: not disclosed.
[m]PST: problem-solving therapy.
[n]bCBT: blended cognitive behavioral therapy.
[o]UTP: unified transdiagnostic protocol.
[p]ACT: acceptance and commitment therapy.
[q]DBT: dialectical behavioral therapy.
[r]BA: behavioral activation.
[s]Not applicable.
[t]FCR: fear of cancer recurrence.

Primary outcomes of individual studies included efficacy or effectiveness (14/29, 48%) [37-39,45,46,48,54,55,57,60,61,63-65]; working alliance (3/29, 10%) [43,53,62]; usability and uptake (3/29, 10%) [50,51,59]; feasibility (2/29, 7%) [23,42]; and patient or therapist (4/29, 14%) [40,41,49,56] experience. One (4%) study [47] explored multiple primary outcomes of therapeutic alliance, compliance, and symptom reduction, and 2 (7%) studies [26,52] reported dual outcomes of efficacy and cost.

**Table 2.** Patient participants' characteristics.

| Study | Age (y), mean (SD); range | Female, n (%) | Pre-post time points, assessment tool, psychological symptoms per group: mean (SD) and severity at baseline |
|---|---|---|---|
| Askjer and Mathiasen [43], 2021 | 35 (13.96); 18-71 | 56 (74) | • Baseline and 12 wk<br>• PHQ-9[a], Depression<br>  • BT[b]: 14.4 (4.1), moderate to severe<br>  • Comparator: 16.05 (3.8), moderate to severe |
| Berger et al [45], 2018 | 43 (12.0); 19-73 | 65 (66) | • Baseline and 12 wk<br>• BDI-II[c], Depression<br>  • BT: 29.6 (8.4), severe<br>  • Comparator: 30.2 (11.2), severe<br>• GAD-7[d], Anxiety<br>  • BT: 11.7 (4.9), moderate<br>  • Comparator: 12 (4.6), moderate |
| Bisson et al [46], 2022 | 36 (13.4); 18 to >65 | 125 (64) | • Baseline and 16 wk<br>• CAPS-5[e], PTSD[f]<br>  • BT: 34.6 (6.8), mild to moderate<br>  • Comparator: 35.6 (6.7), mild-moderate<br>• PHQ-9, Depression<br>  • BT: 15.1 (6.7), moderate to severe<br>  • Comparator: 13.4 (4.6), moderate<br>• GAD-7, Anxiety<br>  • BT: 13.9 (4.9), moderate<br>  • Comparator: 13.4 (4.6), moderate |
| Cloitre et al [47], 2022 | 44 (11.73); 22-77 | 122 (60) | • Baseline; 10 wk<br>• PCL-5[g], PTSD<br>  • BT: 50.7 (15.5), severe<br>• PHQ-9, Depression<br>  • BT: 15.6 (5.4), moderate to severe<br>  • Comparator:—[h] |
| Duffy et al [48], 2020 | 41 (13.1); 17-80 | 85 (69) | • Baseline; at iCBT[i] end<br>• PHQ-9, Depression<br>  • BT: 15.6 (5.5), moderate to severe<br>• GAD-7, Anxiety<br>  • BT: 14.8 (4.5), moderate to severe<br>  • Comparator: — |
| Etzelmueller et al [49], 2018 | 55 (—); 24-64 | 18 (72) | • Weekly assessments, —<br>• QIDS-16-SR[j], Depression<br>  • BT: 14.8 (4.4), moderate<br>  • Comparator: — |
| Høifødt et al [37], 2013 | 36 (11.3); 19-63 | 77 (73) | • Baseline and 7 wk<br>• BDI-II, Depression<br>  • BT: 21.1 (6.85), moderate<br>  • Comparator: 22.3 (6.7), moderate<br>• BAI[k], Anxiety<br>  • BT: 12.05 (11.1), mild<br>  • Comparator: 15.3 (10.9), mild |

| Study | Age (y), mean (SD); range | Female, n (%) | Pre-post time points, assessment tool, psychological symptoms per group: mean (SD) and severity at baseline |
|---|---|---|---|
| Jacmon et al [42], 2009 | 35 (10.37); — | 4 (44) | • Baseline; 6 wk<br>• BDI-II, Depression<br>  • BT: 26.5 (1.5), moderate<br>  • Comparator: — |
| Kemmeren et al [50], 2019 | 42 (12.9); 18-74 | 129 (64) | • Baseline; —<br>• PHQ-9, Depression<br>  • BT: 16.2 (4.7), moderate to severe<br>  • Comparator: — |
| Kenter et al [51], 2013 | 37 (10.8); 18-61 | 73 (70) | • Baseline; 5 wk<br>• BDI-II, Depression<br>  • BT: 23.5 (7), moderate<br>  • Comparator: 22.4 (8.7)<br>• HADS[l], Anxiety<br>  • BT: 10.6 (2.9), moderate<br>  • Comparator: 11.1 (2.4)<br>• MBI[m], Burnout<br>  • MBI-EE, BT: 2.5 (1.7); Comparator: 2.6 (1.4)<br>  • MBI-D, BT: 2.0 (1.5); Comparator: 2.0 (1.5)<br>  • MBI-C, BT: 3.5 (1.3); Comparator: 3.6 (1.3) |
| Kenter et al [52], 2015 | 47 (18.7); 18-91 | 2442 (55) | • First and last F2F sessions<br>• GAF[n], Depression group<br>  • BT: 54.6 (4.9), moderate<br>  • Comparator: 54.7 (4.7), moderate<br>• GAF, Anxiety group<br>  • BT: 59 (5.3), moderate<br>  • Comparator: 59.1 (5.2), moderate |
| Kok et al [38], 2014 | 35 (11.7); — | 130 (61) | • Baseline; 5 wk<br>• FQ[o], Phobia<br>  • BT: 42.4 (23.4)<br>  • Comparator: 38.2 (21.9)<br>• CES-D[p], Depression<br>  • BT: 25 (8.6), severe<br>  • Comparator: 24.7 (8.4), severe<br>• BAI, Anxiety<br>  • BT: 45 (13.8), severe<br>  • Comparator: 44.48 (13.1), severe |
| Kooistra et al [23], 2016 | 38 (8.36); 27-50 | 5 (55) | • Baseline; 10 wk<br>• IDS-SR30[q], Depression:<br>  • BT: 40.4 (12.9), moderate<br>• BAI, Anxiety<br>  • BT: 22 (12.4), moderate<br>  • Comparator: — |
| Kooistra et al [26], 2019 | 39 (10.9); — | 64 (63) | • Baseline; 10 wk<br>• IDS-SR30, Depression:<br>  • BT: 45.2 (12.1), moderate<br>  • Comparator: 41.5 (11.6), moderate |
| Kooistra et al [53], 2020 | 38 (11.0); — | 43 (60) | |

| Study | Age (y), mean (SD); range | Female, n (%) | Pre-post time points, assessment tool, psychological symptoms per group: mean (SD) and severity at baseline |
|---|---|---|---|
| | | | • Baseline; 10 wk<br>• QIDS-SR, Depression<br>  • BT: 16.6 (4.9), severe<br>  • Comparator: 15.9 (4.1), severe |
| Lungu et al [54], 2020 | 33 (8.0); — | 244 (64) | • Baseline and 6 wk<br>• PHQ-9, Depression<br>  • BT: 10.8 (4.7), moderate<br>• GAD-7, Anxiety<br>  • BT: 11.7 (3.9), moderate<br>  • Comparator: — |
| Ly et al [55], 2015 | 31 (11.4); 18-73 | 65 (70) | • Baseline; 9 wk<br>• BDI-II, Depression<br>  • BT: 29 (8.1), severe;<br>  • Comparator: 27.3 (7.9), severe<br>• PHQ-9, Depression<br>  • BT: 15.4 (4.7), moderate to severe;<br>  • Comparator: 15.3 (4.5), moderate to severe<br>• BAI, Anxiety<br>  • BT: 15.7 (12.1), mild to moderate;<br>  • Comparator: 17.5 (9.2), moderate |
| Månsson et al [39], 2013 | 43 (15); 22-70 | 10 (67) | • Baseline; 9 wk<br>• BAI, Anxiety<br>  • BT: 18.1 (7.7), moderate<br>• GAD-7, Anxiety<br>  • BT: 11.9 (6), moderate<br>• PHQ-9, Depression<br>  • BT: 12.1 (6), moderate<br>• MADRS-S$^{r}$, Depression<br>  • BT: 21.2 (4), moderate<br>  • Comparator: — |
| Månsson et al [40], 2017 | 30 (10.6); 18-60 | 36 (80) | • Baseline; 12 wk<br>• PHQ-9, Depression<br>  • BT: 13.3 (5.2)<br>• MADRS-S, Depression<br>  • BT: 21 (8.7), moderate<br>• BAI, Anxiety<br>  • BT: 20.5 (9.8)<br>• GAD-7, Anxiety<br>  • BT: 10.3 (6), moderate<br>  • Comparator: — |
| Mol et al [56], 2018 | 36 (12.3); 21-64 | 33 (73) | • Baseline; approximately 26 wk<br>• QIDS, Depression<br>  • BT: 15.8 (3.8), moderate to severe<br>  • Comparator: — |
| Nakao et al [57], 2018 | 40 (9.7); — | 20 (50) | |

| Study | Age (y), mean (SD); range | Female, n (%) | Pre-post time points, assessment tool, psychological symptoms per group: mean (SD) and severity at baseline |
|---|---|---|---|
| | | | • Baseline; 12 wk<br>• GRID-HAMD[s], Depression<br>  • BT: 18.3 (3.7), moderate to severe;<br>  • Comparator: 18.5 (3.6), moderate to severe<br>• BDI-II, Depression<br>  • BT: 28 (8.8), moderate;<br>  • Comparator: 24.4 (7.8), moderate<br>• QIDS, Depression<br>  • BT: 14.8 (4.2), moderate;<br>  • Comparator: 13.5 (4), moderate |
| Romijn et al [41], 2021 | 37 (11.0); 19-62 | 23 (52) | • Baseline; 15 wk<br>• BDI-II, Depression<br>  • BT: 24 (12.2), moderate;<br>  • Comparator: 24 (10.3), moderate<br>• BAI, Anxiety<br>  • BT: 27.9 (12), severe;<br>  • Comparator: 27.15 (11.7), severe |
| Tarp et al [59], 2022 | 47 (12); 28-73 | 7 (32) | • Addictive disorder: — |
| Thase et al [60], 2018 | 46 (14.3); — | 102 (66) | • Baseline and week 16<br>• Depression: HDRS[t]<br>  • BT: 19.8 (3.5), moderate to severe<br>  • Comparator: 19.6 (3.8), moderate to severe |
| van de Wal et al [61], 2017 | 59 (11.3); 31-77 | 47 (53) | • Baseline and 12 wk<br>• CWS[u], FCR<br>  • BT: 19.6 (3.7), high;<br>  • Comparator: 19.6 (3.7), high<br>• HADS-D, Depression<br>  • BT: 5.9 (4.2), low;<br>  • Comparator: 6.8 (4.7), low<br>• HADS-A, Anxiety<br>  • BT: 8.1 (4.1), mild;<br>  • Comparator: 8.4 (4.9), mild |
| Vernmark et al [62], 2019 | 35 (13.9); — | 54 (74) | • Baseline; 10 wk<br>• PHQ-9, Depression<br>  • BT: 14.3 (5.1), moderate<br>  • Comparator: — |
| Witlox et al [63], 2021 | 63 (5.70); 55-75 | 192 (61) | • Baseline; 12 wk<br>• PHQ-9, Depression<br>  • BT: 7 (4), mild;<br>  • Comparator: 7.9 (3.5)<br>• GAD-7, Anxiety<br>  • BT: 8.2 (4), mild to moderate;<br>  • Comparator: 8.8 (3) |
| Wu et al [64], 2021 | 33 (8.62); — | 921 (62) | |

| Study | Age (y), mean (SD); range | Female, n (%) | Pre-post time points, assessment tool, psychological symptoms per group: mean (SD) and severity at baseline |
|---|---|---|---|
| | | | • Wk 0-7 and wk 8-15<br>• PHQ-9, Depression<br>  • BT: 6 (3), mild<br>• GAD-7, Anxiety<br>  • BT: 9.6 (2), mild to moderate<br>  • Comparator: — |
| Wu et al [65], 2021 | 33 (8.68); — | 2218 (65) | • Wk 0-7 and wk 8-15<br>• PHQ-9, Depression<br>  • BT: 10.7 (4.9), moderate<br>• GAD-7, Anxiety<br>  • BT: 11.8 (4.1), moderate<br>  • Comparator: — |

[a]PHQ-9: Patient Health Questionnaire-9.

[b]BT: blended therapy.

[c]BDI-II: Beck Depression Inventory II.

[d]GAD-7: Generalized Anxiety Disorder-7.

[e]CAPS-5: Clinician-Administered Posttraumatic Stress Disorder Scale for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[f]PTSD: posttraumatic stress disorder.

[g]PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[h]Not available.

[i]iCBT: internet-delivered cognitive behavioral therapy.

[j]QIDS-16-SR: 16-Item Quick Inventory of Depressive Symptomatology, self-reported.

[k]BAI: Beck Anxiety Inventory.

[l]HADS: Hospital Anxiety and Depression Scale.

[m]MBI: Maslach Burnout Inventory.

[n]GAF: Global Assessment of Functioning.

[o]FQ: Fear Questionnaire.

[p]CES-D: Centre for Epidemiologic Studies Depression Scale.

[q]IDS-SR30: Inventory of Depressive Symptomatology Self-Rated.

[r]MADRS-S: Montgomery Åsberg Depression Rating Scale—self-rating version.

[s]GRID-HAMD: 17-item GRID-Hamilton Depression Rating Scale score.

[t]HDRS: Hamilton Depression Rating Scale.

[u]CWS: Cancer Worry Scale.

## Classification of BT Models

On the basis of the models defined by Erbe et al [24], the majority (26/29, 90%) of the studies reported an *integrated* intervention design, where new therapeutic content was delivered either (1) across both the face-to-face and digital modalities (*core*; n=14) or (2) primarily using one modality (usually the face-to-face component) with additional content delivered as *supplementary material* (n=12). Integrated interventions were the most common designs for addressing depression (n=14) and depression and or anxiety (n=8). Three studies were classified as *sequential* designs and described a *core* role for both face-to-face and digital components. One sequential model was delivered as *stepped care* in which most patients received digital therapy only.

BT delivery was further differentiated based on whether both components were delivered in a preset, *alternate* format or a tailored, *case-by-case* arrangement. Most integrated interventions (17/26, 65%) used an *alternate* delivery format. Therapeutic content within integrated and alternate designs had either *core* (9/17, 53%) or *supplementary* (8/17, 47%) roles.

In 19 (65%) studies, patients engaged with digital content following a *standardized* program with minimal tailoring of the materials presented. In contrast, 10 (35%) studies described a more *personalized* manner of assigning or interacting with digital content where the therapist and patient have more autonomy to choose, change, or create digital content according to individual need. An overview of model classifications is presented in Table 3.

**Table 3.** Blended therapy model classification per study.

| Study | Interaction of F2F[a] and digital components | | Role of F2F and digital components | | Pattern of delivery of F2F and digital | | Digital content delivery | |
|---|---|---|---|---|---|---|---|---|
| | Integrated | Sequential | Core | Supplementary | Alternate | Case by case | Personalized | Standardized |
| Askjer and Mathiasen [43], 2021 | ✓ | | ✓ | | ✓ | | | ✓ |
| Berger et al [45], 2018 | ✓ | | | ✓ | | ✓ | ✓ | |
| Bisson et al [46], 2022 | ✓ | | | ✓ | | ✓ | ✓ | |
| Cloitre et al [47], 2022 | ✓ | | | ✓ | ✓ | | | ✓ |
| Duffy et al [48], 2020[b] | | ✓ | ✓ | | | | ✓ | |
| Etzelmueller et al [49], 2018 | ✓ | | | ✓ | | ✓ | ✓ | |
| Høifødt et al [37], 2013 | ✓ | | ✓ | | ✓ | | | ✓ |
| Jacmon et al [42], 2009 | ✓ | | ✓ | | | ✓ | | ✓ |
| Kemmeren et al [50], 2019 | ✓ | | ✓ | | ✓ | | | ✓ |
| Kenter et al [51], 2013[b] | | ✓ | ✓ | | | | | ✓ |
| Kenter et al [52], 2015 | ✓ | | ✓ | | | ✓ | | ✓ |
| Kok et al [38], 2014[b] | | ✓ | ✓ | | | | | ✓ |
| Kooistra et al [23], 2016 | ✓ | | | ✓ | ✓ | | | ✓ |
| Kooistra et al [26], 2019 | ✓ | | | ✓ | ✓ | | | ✓ |
| Kooistra et al [53], 2020 | ✓ | | | ✓ | ✓ | | | ✓ |
| Lungu et al [54], 2020 | ✓ | | | ✓ | ✓ | | ✓ | |
| Ly et al [55], 2015 | ✓ | | | ✓ | ✓ | | ✓ | |
| Månsson et al [39], 2013 | ✓ | | | ✓ | | ✓ | ✓ | |
| Månsson et al [40], 2017 | ✓ | | | ✓ | | ✓ | ✓ | |
| Mol et al [56], 2018 | ✓ | | ✓ | | ✓ | | | ✓ |
| Nakao et al [57], 2018 | ✓ | | ✓ | | ✓ | | ✓ | |
| Romijn G et al [41], 2021 | ✓ | | ✓ | | ✓ | | | ✓ |
| Tarp et al [59], 2022 | ✓ | | ✓ | | | ✓ | | ✓ |
| Thase et al [60], 2018 | ✓ | | ✓ | | ✓ | | | ✓ |
| van de Wal et al [61], 2017 | ✓ | | | ✓ | ✓ | | | ✓ |

| Study | Interaction of F2F[a] and digital components | | Role of F2F and digital components | | Pattern of delivery of F2F and digital | | Digital content delivery | |
|---|---|---|---|---|---|---|---|---|
| | Integrated | Sequential | Core | Supplementary | Alternate | Case by case | Personalized | Standardized |
| Vernmark et al [62], 2019 | ✓ | | ✓ | | ✓ | | | ✓ |
| Witlox et al [63], 2021 | ✓ | | ✓ | | ✓ | | | ✓ |
| Wu et al [64], 2021 | ✓ | | ✓ | | | ✓ | ✓ | |
| Wu et al [65], 2021 | ✓ | | ✓ | | | ✓ | ✓ | |

[a]F2F: face to face.

[b]Sequential models present a *linear* pattern of delivery.

## Structure, Content, and Ratio of Sessions in BT Models

### *Overview*

Within the *integrated* model, 81% (21/26) of studies reported commencing BT with the face-to-face component. Five *integrated* interventions began treatment with the digital component; of those, 4 [38,42,47,51] used the digital component as the intervention "anchor"—that is, the digital component led the therapeutic process. One [48] *sequential* intervention (stepped care) used the digital modality as a prequel for high-intensity face-to-face treatment based on patient symptom severity. Digital sessions were mostly delivered via website platforms with individualized access. Overall, digital session components presented CBT-based content and followed the frameworks used in face-to-face settings. Digital content was typically asynchronous. Intervention structure and content is summarized in Multimedia Appendix 4.

### *Face-to-Face Versus Digital Sessions Distribution in BT Models*

In total, 26 (87%) studies reported the number of face-to-face sessions, which ranged from 3 to 21 sessions (mean 7, SD 4.2). Of those reporting the number of face-to-face sessions, 24 (92%) used an *integrated* intervention design. Mean face-to-face session duration across studies was 49 minutes (SD 11.7, range 27-65 min/session). Periodicity of sessions were reported by 24 studies, with most of those studies describing face-to-face sessions as delivered weekly (12/24, 50%) [23,26,37,39,41,42,45,49,53,57,59,64].

A total of 20 (69%) studies reported on the number of digital sessions (mean 8, SD 3, range 4-14 sessions)—of those studies, 18 (90%) presented an *integrated* design. Although the mean time for digital sessions was largely undefined and unreported, 12 (41%) studies [37,40,45,47,50,52,57,60-63,65] reported that digital modules were typically developed to range between 15 and 60 minutes. Eight (28%) studies [42,45,46,48,57,59,60,64] allowed patients to complete modules at their own pace; 13 (45%) studies [39,40,42,45,48,49,54-56,60,61,64,65] reported digital sessions' periodicity to be flexible; weekly (14/29, 48%) [23,26,37,38,41,46,47,51-53,57,59,62,63] or fortnightly (2/29, 7%) [43,50] completion was also reported.

### *Face-to-Face and Digital Ratio*

In total, 16 (55%) studies [23,26,37,41,42,45-50,53,56,60,62,65] reported the ratio of face-to-face and digital sessions. Most of those studies had an *integrated* design (14/16, 88%) and typically delivered sessions at a 1:1 ratio (1 face-to-face to 1 digital).

### *Treatment Length*

The BT mean length of treatment was 12 (SD 5.1, range 6-26) weeks; however, this was unreported for 5 studies [38,49,51,52,63] (Table 4).

**Table 4.** Model classification versus session structure.

| | Integrated (n=26) | Sequential (n=3) | Core (n=17) | Supplementary (n=12) | Alternate (n=17) | Case by case (n=9) | Linear[a] (n=3) | Personalized (n=10) | Standardized (n=19) |
|---|---|---|---|---|---|---|---|---|---|
| Treatment length[b], mean (range); n | 12 (6-21); 23 | 21; (—[c]); 1 | 13 (6-26); 13 | 11 (6-21); 11 | 13 (6-26); 16 | 10 (6-12); 7 | 21; (—); 1 | 10 (6-21); 9 | 14 (8-26); 15 |
| **Face-to-face sessions** | | | | | | | | | |
| Number, mean (range); n | 8 (3.5-21); 23 | 3.5 (0-10); 3 | 6 (0-12); 15 | 9 (4-21); 11 | 7 (3.5-12); 17 | 9 (4-21); 6 | 10; (—); 1 | 9 (4-21); 9 | 6 (0-12); 17 |
| Time (min), mean (range); n | 49 (27-65); 17 | NR[d]; 0 | 47 (27-65); 9 | 51 (45-65); 8 | 46.5 (27-65); 12 | 55 (45-60); 5 | NR; 0 | 53 (44-60); 6 | 47 (27-65); 11 |
| Periodicity | 1w[e]: n=12; 2w[f]: n=4; 3w[g]: n=1; var[h]: n=7 | NR; (0) | 1w: n=6; 2w: n=3; 3w: n=1; var: n=3 | 1w: n=6; var: n=4; 2w: n=1 | 1w: n=6; 2w: n=4; 3w: n=1; var: n=6 | 1w: n=6; var: n=1 | NR | 1w: n=5; var: n=3 | 1w: n=7; 2w: n=4; 3w: n=1; var: n=4 |
| **Digital sessions** | | | | | | | | | |
| Number, mean (range); n | 9 (4-14); 17 | 4 (4-5); 2 | 7 (4-14); 11 | 9 (7-14); 8 | 7 (4-10); 11 | 10 (6-14); 6 | 4; 1 | 8 (4-14); 5 | 8 (4-14); 15 |
| Time (min), mean (range); n | 38 (6-65); 8 | NR; 0 | 44 (30-65); 4 | 32 (6-62); 4 | 38 (28-65); 5 | 38 (6-62); 3 | NR; 0 | 36 (6-62); 4 | 40 (28-65); 4 |
| Periodicity | 1w: n=13; 2w: n=1; var: n=12 | 2w: n=1; any[i]: n=2 | 1w n=9; 2w: n=2; any: n=6 | 2w: n=4; any: n=8 | 1w: n=9; 2w: n=2; any: n=6 | 1w: n=2; any: n=7 | 1w: n=2; any: n=1 | 1w: n=1; any: n=9 | 1w: n=12; 2w; n=2; any: n=5 |

[a]*Linear* describes the delivery pattern typical of *sequential* designs.

[b]Treatment length is measured in weeks.

[c]Not applicable.

[d]NR: not reported.

[e]1w: weekly.

[f]2w: 2-weekly.

[g]3w: 3-weekly.

[h]var: variable (eg, varied from weekly to fortnightly or other pattern).

[i]any: anytime (ie, no pattern from the outset).

## Outcomes of BT Interventions

### *Treatment Uptake*

In total, 25 (86%) studies reported on BT uptake mean rates (mean 91%, SD 10.2%, range 60%-100%). Of those, 24 (96%) studies [23,26,37-43,45-48,50,51,53,55-57,59-63] reported a mean uptake of 92% (SD 8.2%, range 73%-100%) and 1 (4%) study [49] reported a lower uptake (60%). A total of 14 (48%) studies reported on treatment completion rates, with an average of 61% (SD 29.4%, range 11.5%-100%) of patients completing treatment.

### *Treatment Adherence*

Our review considered adherence as completing a minimum number of sessions determined by each study. However, adherence criteria differed across studies and there was a lack of data to confirm whether session structure or ratio influenced adherence to intervention. A total of 23 (279%) studies reported on BT adherence rates, with a mean of 81% (SD 11.8%). In total, 20 (69%) *integrated* studies reported on BT adherence (mean 83%, SD 11%, range 62%-100%). In contrast, all *sequential* studies (3/29, 10%) had a comparatively lower mean adherence (64%, SD 1.8). Dropout rates were reported in 25 (86%) studies, with 24 of those reporting <40% dropout rates (mean 18.5%, SD 17.2%, range 0%-38%). The intervention with the lowest adherence rate (16%) and highest dropout rate (84%) [38] had a *sequential* design.

Regarding the role of components within BT, interventions with *supplementary* designs presented higher adherence (mean 88%, SD 9%, range 72.5%-100%) than *core* designs (mean 76%, SD 11.7%, range 16%-100%). Details on the uptake and adherence reported per study are available in Multimedia Appendix 5.

### *Health Service Outcomes*

Four (14%) studies reported on the impact of BT on therapist time: 2 studies [42,55] found a decrease, one study reported an increase [52], and one found no change [26] in the time needed to deliver therapy. Findings related to efficacy and costs were mixed, with one study [52] suggesting that BT was not cost-effective compared with face-to-face therapy, while another study [55] highlighted reduced costs due to the potential of BT for treating twice as many patients as compared with face-to-face treatment.

### *Patient Satisfaction and Working Alliance*

Eight (28%) studies [23,37,38,43,45,46,49,63] reported on patient satisfaction with BT treatment—criteria for determining patient satisfaction were heterogeneous, but all studies reported it as "high." Four (14%) studies [43,47,53,62] reported on working alliance and reported it as "high." Two (7%) studies [43,62] suggested that the therapist-rated working alliance predicted treatment outcomes and that this may be specific to BT.

### *Barriers to and Facilitators of Intervention Uptake and Engagement*

Potential barriers to intervention uptake reported [23,46,49,51,59] included *a lack of understanding about the intervention* and *digital challenges*. Reported [42,51,59] facilitators to intervention uptake included *convenience and flexibility of the digital component, anonymity,* and *autonomy enabled by the BT design.* Barriers to intervention engagement identified [37,38,42,45,46,48-50,56,61] included *the good enough effect* (ie, when patients drop out during the initial stages of the intervention arguing they "feel better" and "no longer need therapy" [66]), *being left unchecked, reduced therapy support* (ie, lack of therapist follow-up regarding digital activities), and *digital challenges*. Facilitators of intervention engagement included *experience in the use of technology, a program tailored to patient-specific needs, patient-therapist digital communication between sessions,* and *a patient-therapist working alliance fostered throughout the intervention* [37-39,43,45-50,52,57,64,65].

### *Treatment Efficacy*

We conducted a meta-analysis to examine differences in treatment outcomes. There were only a sufficient number of studies to meta-analyze depression and anxiety outcomes.

Data pooled from 9 RCT studies demonstrated a moderate to large, significant improvement in depression symptoms (Cohen *d*=–1.1, 95% CI –0.6 to –1.6, *P*<.001). However, comparing treatment outcomes for anxiety interventions with controls (n=5), there was no significant improvement across studies (Cohen *d*=–0.1, 95% CI –0.3 to 0.05, *P*=.17). Between-study heterogeneity was high and did not change after conducting publication bias assessment using the funnel plot trim and fill method, both in the depression (*Q*=90.3; *P*<.001; $I^2$=91.1; $T^2$=0.1; *T*=0.7) and anxiety groups (*Q*=1.1; *P*=.17; $I^2$=0; $T^2$=0; *T*=0). We noted that estimates of heterogeneity are impacted by the very small number of studies analyzed (<10 studies).

Meta-analysis was also conducted to examine associations between depression and anxiety scores on various scales and BT structure. For depression, mixed-effect analysis suggested higher effects sizes for interventions where the therapeutic content was delivered primarily face to face with digital content *supplementing* the face-to-face content (n=11; Cohen *d*=–0.75, 95% CI –0.56 to –0.95) compared with digital therapeutic content delivered as *core* (n=10; Cohen *d*=–0.5, 95% CI –0.4 to –0.7)—however, differences were not statistically significant (*P*<.08). Similar associations were found for anxiety, with higher ESs for *supplementary* (n=8; Cohen *d*=–0.9, 95% CI –0.6 to –1.2) compared with *core* structure (n=5; Cohen *d*=–0.6, 95% CI –0.22 to –0.98). Mixed-effect analysis also indicated statistically significantly (*P*<.001) higher ESs for interventions with ≤6 face-to-face sessions both for depression (n=9; Cohen *d*=–0.7, 95% CI –0.5 to –0.9) and anxiety (n=7; Cohen *d*=–0.8, 95% CI –0.3 to –1.3) compared with interventions with >6 face-to-face sessions for depression (n=11; Cohen *d*=–0.6, 95% CI –0.4 to –0.8) and anxiety (n=5; Cohen *d*=–0.7, 95% CI –0.4 to –1). Similarly, interventions with >50% of sessions delivered digitally had higher ESs (*P*<.001) both for depression (n=5; Cohen *d*=–0.8, 95% CI –0.6 to –1.1) and anxiety (n=4; Cohen *d*=–0.8, 95% CI 0.006 to –1.6) compared with interventions where >50% sessions were delivered face to face, both for depression (n=9; Cohen *d*=–0.5, 95% CI –0.3 to –0.7) and anxiety (n=3; Cohen *d*=–0.58, 95% CI –0.3 to –0.9). Data regarding meta-analyses are available in Multimedia Appendix 6.

## Discussion

### Overview

We reviewed blended psychological therapy models and classified them according to their structure, content, and ratio of face-to-face and digital sessions. Most BT interventions were CBT-based and addressed depression—for which models with *integrated* and *supplementary* designs resulted in improved treatment efficacy. Interventions typically used an *integrated* design with the face-to-face component "anchoring" the intervention. Essential, therapeutic content across treatment designs was typically delivered as both face to face and digital (ie, a *core* design) and in an *alternate* pattern. However, several studies used digital components to *supplement* therapeutic content delivered face to face. Most interventions also relied on *standardized* digital content rather than content tailored to individual patients.

Our study confirms that BT leads to improved overall patient uptake (mean 91%) and adherence (mean 81%), contrasting with the lower uptake and adherence rates previously reported for digital therapy alone. For example, an observational cohort study analyzed clinical data from 15,882 patients assessed for digital-only treatment of various psychological disorders, reporting 22% uptake and 68% adherence rates [14]. In addition, review studies on digital therapy for depression and anxiety symptoms [16] and for depression alone [17] indicated mean uptake rates of 56% (range 21%-88%) and 88% (range 42%-100%), respectively, as well as mean adherence rates of 18% (range 7%-42%) and 60% (range 14%-93%), respectively.

These contrasting findings suggest a potential connection between improved engagement with digital components when these are integrated into an intervention with a more prominent role of the therapist, that is, in a blended format.

### Meta-Analysis Results: Treatment Versus Control Dyads

Across CBT-based RCTs, symptom reduction was observed for both depression and anxiety—although the reduction in anxiety symptoms was not substantial. Our result contrasts with overall findings on effectiveness of digital-only interventions exclusively addressing anxiety [67], which indicate effective outcomes. However, a subgroup analysis in that study confirmed similar ESs of both digital and face-to-face anxiety treatments. The lack of significance of our results might reflect the fact that only 1 included study addressed anxiety as a primary intervention outcome, while the other 4 interventions primarily targeted posttraumatic stress disorder, Fear of Cancer Recurrence, or depression. This suggests that transdiagnostic interventions may not be sufficient for treating anxiety—even with therapist guidance.

### Meta-Analysis Results: BT Treatment Outcomes Versus the BT Model

Our analysis indicated that interventions that delivered face-to-face sessions for depression and anxiety in a lower ratio (≤50%) and in a lower number (≤6) had higher, statistically significant overall ESs. As the ratio of face-to-face sessions was similar in *core* or *supplementary* models, it is unclear what role digital sessions played in the delivery of therapeutic content compared with simply supplementing face-to-face sessions. Moreover, ESs for depression and anxiety in *supplementary* designs (Cohen *d*=–0.75 and Cohen *d*=–0.87, respectively) were higher than those in *core* designs (Cohen *d*=–0.53 and Cohen *d*=–0.6, respectively). Those findings suggest that therapeutic content may achieve better results when introduced by therapists and reinforced digitally—a characteristic of *supplementary* models, in which the digital component extends or reinforces face-to-face content. In addition, s*upplementary* models might provide a more seamless transition between face-to-face and digital content as therapists can identify and discuss challenging topics with patients before they go on the web, what could result in enhanced patient engagement, adherence and treatment results. This argument is supported by studies on participants' preference for BT models that enable greater therapist-patient interaction [21,50,68,69]. In addition, participants' views reported across our study also suggest that improved digital access and support from therapist facilitated BT engagement.

The higher ESs of *supplementary* versus *core* designs may also reflect the therapists' preference for face-to-face contact. This argument finds support in studies on health professionals' preferences regarding face-to-face versus digital delivery, both in blended [68,70] and digital-only [20,71] interventions. Those studies suggest that, despite recognizing the advantages of digital interventions, professionals perceive face-to-face delivery as more attractive than digital delivery, which could influence the endorsement of digital therapy delivery as a supplement. This suggests that therapists might feel more comfortable *engaging with* and *promoting* the digital arm in a *supplementary* way.

In addition, it is possible that both therapists and patients expect the "bulk" of the therapeutic content to be delivered face to face in blended interventions, explaining why we found BT *core* models to be less efficacious than *supplementary*. Therapy expectations would also help explain the contrast of our results with research [67,72] on using digital therapy alone for depression and anxiety, which found digital interventions to be as good as or better than face-to-face interventions. Perhaps because patients engaging in digital-only treatment would *expect* therapeutic content to be delivered digitally, as there is no face-to-face option, they fare better on the digital component than they would in a BT model.

### Optimizing Effectiveness, Time, and Resources in BT Treatment

One aim of BT interventions [24] is to enable better balance between treatment effects; patient engagement; treatment time; and the use of resources for both patients and therapists. BT models with *integrated*, *core*, *alternate*, and *standardized* designs allow for optimized treatment delivery. However, the results of our meta-analysis suggest that *supplementary* BT models are more effective. Considering these arguments, perhaps a midway alternative would be to promote a more digitally focused role of the therapist in a blended design. Following this idea, therapists would support and encourage patients to complete digital content as an enhancing adjunct to face-to-face contact—and not only as an "add-on" feature. A more digitally focused role of the therapist might promote increased patient engagement, enabling improved outcomes with lower doses of face-to-face treatment.

### Study Limitations

We optimized database search terms with the assistance of a librarian; however, it is possible that eligible studies were not included in this review. In addition, the analysis did not assess the potential impact of heterogeneities found among selected studies on the variables analyzed, for example, whether different populations, settings, or therapeutic approach applied in the interventions might affect uptake or adherence or even the way blended sessions are delivered. Furthermore, the small number of studies included in our meta-analysis impacted the breadth of our results. In addition, there were limited studies describing intervention structure and content of both face-to-face and digital components in detail. Despite having contacted authors regarding missing data, several gaps remained on key variables investigated (eg, number and time of face-to-face and digital sessions, adherence parameters, and acceptability of intervention), which impacted the depth of analysis. Hence, the outcomes reported in this study should be interpreted with caution.

### Recommendations

Despite growing evidence regarding BT efficacy, the lack of clearer, detailed data reporting on its structure and content poses challenges to scientific reproducibility of BT, possibly affecting its implementation success. Details related to the number, time, and distribution of sessions; specific content of both digital and face-to-face sessions, as well as the role, relevance, and influence of the digital component within the therapy plan;

feedback on face-to-face and digital session content; and the use of assigned digital materials as well as the perception of its usefulness from the perspective of both therapists and patients are examples of useful data that are not adequately reported. We recommend that future studies include a more detailed reporting of methodology, particularly regarding the structure and content of sessions.

### Conclusions

This systematic review examined blended models for the treatment of psychological disorders to identify what aspects of BT underpin effective treatment and improved engagement. Evidence suggests that implementing an integrated model is feasible in the treatment of psychological disorders. BT was reported as being either more effective or noninferior to face-to-face treatment, particularly when applied to the treatment of anxiety and depression. BT interventions studied reported high mean uptake and adherence rates, showing promise in improving engagement to treatment. Higher ESs were found for depression and anxiety outcomes in interventions with *integrated*, *supplementary* models; with a lower number of face-to-face sessions; and with a lower ratio of face-to-face versus digital sessions, suggesting that combining a more digitally focused therapist role with fewer face-to-face sessions can be effective and increase access to treatment.

To support improved reporting, we have developed a taxonomy for BT models based on the key themes identified in this review regarding model structure and components. Future studies detailing the structure and content of BT models may help identify suitable models for the treatment of different psychological disorders.

### Acknowledgments

The Psycho-oncology Co-operative Research Group supported the development of this research. The Psycho-oncology Co-operative Research Group is funded by Cancer Australia through their Support for Clinical Trials Funding Scheme.

### Authors' Contributions

KFN-Z, HLS, PB, LB, and JMS were responsible for conceptualization and protocol development. The literature search was conducted by KFN-Z and JMS, and study selection was carried out by KFN-Z, JMS, and PB. Data analysis was performed by KFN-Z and JMS. The initial draft was prepared by KFN-Z and JMS, and all authors (KFN-Z, HS, PB, LB, and JMS) participated in reviewing and editing the draft.

### Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 158 KB - ijmr_v13i1e49660_app1.pdf ]

Multimedia Appendix 2
Database searches: examples.
[PDF File (Adobe PDF File), 382 KB - ijmr_v13i1e49660_app2.pdf ]

Multimedia Appendix 3
Quality assessment.
[PDF File (Adobe PDF File), 198 KB - ijmr_v13i1e49660_app3.pdf ]

Multimedia Appendix 4
Blended therapy structure: content described.
[PDF File (Adobe PDF File), 330 KB - ijmr_v13i1e49660_app4.pdf ]

Multimedia Appendix 5
Uptake: adherence described.
[PDF File (Adobe PDF File), 176 KB - ijmr_v13i1e49660_app5.pdf ]

Multimedia Appendix 6
Meta-analyses data: graphs.
[PDF File (Adobe PDF File), 619 KB - ijmr_v13i1e49660_app6.pdf ]

### References

1. Comprehensive mental health action plan 2013-2030. World Health Organization (WHO). 2021. URL: https://www.who.int/publications/i/item/9789240031029 [accessed 2024-04-29]
2. World health statistics 2022: monitoring health for the SDGs, sustainable development goals. World Health Organization (WHO). 2022. URL: https://www.who.int/publications/i/item/9789240051157 [accessed 2024-04-29]
3. The WHO special initiative for mental health (2019-2023): universal health coverage for mental health. World Health Organization (WHO). 2019. URL: https://iris.who.int/handle/10665/310981 [accessed 2024-04-29]
4. Guzman D, Ann-Yi S, Bruera E, Wu J, Williams JL, Najera J, et al. Enhancing palliative care patient access to psychological counseling through outreach telehealth services. Psychooncology 2020 Jan 10;29(1):132-138. [doi: 10.1002/pon.5270] [Medline: 31707735]
5. Sinclair C, Holloway K, Riley G, Auret K. Online mental health resources in rural Australia: clinician perceptions of acceptability. J Med Internet Res 2013 Sep;15(9):e193 [FREE Full text] [doi: 10.2196/jmir.2772] [Medline: 24007949]
6. Hadjistavropoulos HD, Schneider LH, Hadjistavropoulos T, Titov N, Dear BF. Effectiveness, acceptability and feasibility of an internet-delivered cognitive behavioral pain management program in a routine online therapy clinic in Canada. Can J Pain 2018 Mar 21;2(1):62-73 [FREE Full text] [doi: 10.1080/24740527.2018.1442675] [Medline: 35005367]
7. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
8. Christensen H, Batterham P, Calear A. Online interventions for anxiety disorders. Curr Opin Psychiatry 2014 Jan;27(1):7-13. [doi: 10.1097/YCO.0000000000000019] [Medline: 24257123]
9. Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G, Kessler D. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial. Br J Psychiatry 2010 Oct 02;197(4):297-304. [doi: 10.1192/bjp.bp.109.073080] [Medline: 20884953]
10. Saddichha S, Al-Desouki M, Lamia A, Linden IA, Krausz M. Online interventions for depression and anxiety - a systematic review. Health Psychol Behav Med 2014 Jan 01;2(1):841-881 [FREE Full text] [doi: 10.1080/21642850.2014.945934] [Medline: 25750823]
11. Andersson G, Topooco N, Havik O, Nordgreen T. Internet-supported versus face-to-face cognitive behavior therapy for depression. Expert Rev Neurother 2016 Dec 15;16(1):55-60. [doi: 10.1586/14737175.2015.1125783] [Medline: 26610160]
12. Emmelkamp PM. Effectiveness of cybertherapy in mental health: a critical appraisal. Stud Health Technol Inform 2011;167:3-8. [Medline: 21685633]
13. Bisby MA, Karin E, Hathway T, Scott AJ, Heriseanu AI, Dudeney J, et al. A meta-analytic review of randomized clinical trials of online treatments for anxiety: inclusion/exclusion criteria, uptake, adherence, dropout, and clinical outcomes. J Anxiety Disord 2022 Dec;92:102638. [doi: 10.1016/j.janxdis.2022.102638] [Medline: 36242790]
14. Cross SP, Karin E, Staples LG, Bisby MA, Ryan K, Duke G, et al. Factors associated with treatment uptake, completion, and subsequent symptom improvement in a national digital mental health service. Internet Interv 2022 Mar;27:100506 [FREE Full text] [doi: 10.1016/j.invent.2022.100506] [Medline: 35242587]
15. Beatty L, Kemp E, Koczwara B. Finding my way from clinical trial to open access dissemination: comparison of uptake, adherence, and psychosocial outcomes of an online program for cancer-related distress. Support Care Cancer 2022 Jun 22;30(10):7935-7942. [doi: 10.1007/S00520-022-07205-0]
16. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
17. Lipschitz JM, Van Boxtel R, Torous J, Firth J, Lebovitz JG, Burdick KE, et al. Digital mental health interventions for depression: scoping review of user engagement. J Med Internet Res 2022 Oct 14;24(10):e39204 [FREE Full text] [doi: 10.2196/39204] [Medline: 36240001]
18. Klein B, Cook S. Preferences for e-mental health services amongst an online Australian sample? E J Appl Psychol 2010 May 25;6(1):28-39. [doi: 10.7790/ejap.v6i1.184]
19. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
20. Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J. Implementing web-based therapy in routine mental health care: systematic review of health professionals' perspectives. J Med Internet Res 2020 Jul 23;22(7):e17362 [FREE Full text] [doi: 10.2196/17362] [Medline: 32706713]
21. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355. [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
22. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]

23. Kooistra LC, Ruwaard J, Wiersma JE, van Oppen P, van der Vaart R, van Gemert-Pijnen JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. Internet Interv 2016 May;4:61-71 [FREE Full text] [doi: 10.1016/j.invent.2016.01.003] [Medline: 30135791]
24. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
25. Fitzpatrick M, Nedeljkovic M, Abbott JA, Kyrios M, Moulding R. "Blended" therapy: the development and pilot evaluation of an internet-facilitated cognitive behavioral intervention to supplement face-to-face therapy for hoarding disorder. Internet Interv 2018 Jun;12:16-25 [FREE Full text] [doi: 10.1016/j.invent.2018.02.006] [Medline: 30135765]
26. Kooistra LC, Wiersma JE, Ruwaard J, Neijenhuijs K, Lokkerbol J, van Oppen P, et al. Cost and effectiveness of blended versus standard cognitive behavioral therapy for outpatients with depression in routine specialized mental health care: pilot randomized controlled trial. J Med Internet Res 2019 Oct 29;21(10):e14261 [FREE Full text] [doi: 10.2196/14261] [Medline: 31663855]
27. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative effectiveness research on blended depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Aug 03;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
29. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2024-04-29]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Neuendorf KA. The Content Analysis Guidebook. 2nd edition. Thousand Oaks, CA: Sage Publications; 2017.
32. Comprehensive meta-analysis. Version 4. Biostat, Inc. 2022. URL: https://meta-analysis.com/ [accessed 2024-04-29]
33. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
34. Study quality assessment tools. The National Heart, Lung, and Blood Institute (NHLBI). 2021. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2024-04-29]
35. CASP checklist: CASP qualitative studies checklist. Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/qualitative-studies-checklist/ [accessed 2024-04-29]
36. Hong Q, Pluye P, Fàbregues S, Bartlett G, Boardman F. Mixed methods appraisal tool (MMAT) version 2018. Department of Family Medicine. 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf [accessed 2024-04-29]
37. Høifødt RS, Lillevoll KR, Griffiths KM, Wilsgaard T, Eisemann M, Waterloo K, et al. The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: randomized controlled trial. J Med Internet Res 2013 Aug;15(8):e153 [FREE Full text] [doi: 10.2196/jmir.2714] [Medline: 23916965]
38. Kok RN, van Straten A, Beekman AT, Cuijpers P. Short-term effectiveness of web-based guided self-help for phobic outpatients: randomized controlled trial. J Med Internet Res 2014 Sep 29;16(9):e226 [FREE Full text] [doi: 10.2196/jmir.3429] [Medline: 25266929]
39. Månsson KN, Skagius RE, Gervind E, Dahlin M, Andersson G. Development and initial evaluation of an internet-based support system for face-to-face cognitive behavior therapy: a proof of concept study. J Med Internet Res 2013 Dec 10;15(12):e280 [FREE Full text] [doi: 10.2196/jmir.3031] [Medline: 24326065]
40. Månsson KN, Klintmalm H, Nordqvist R, Andersson G. Conventional cognitive behavioral therapy facilitated by an internet-based support system: feasibility study at a psychiatric outpatient clinic. JMIR Res Protoc 2017 Aug 24;6(8):e158 [FREE Full text] [doi: 10.2196/resprot.6035] [Medline: 28838884]
41. Romijn G, Provoost S, Batelaan N, Koning J, van Balkom A, Riper H. Does it blend? Exploring therapist fidelity in blended CBT for anxiety disorders. Internet Interv 2021 Sep;25:100418 [FREE Full text] [doi: 10.1016/j.invent.2021.100418] [Medline: 34401377]
42. Jacmon J, Malouff JM, Taylor N. Treatment of major depression: effectiveness of cognitive behavior therapy with an internet course as a central component. E J Appl Psychol 2009;5(2):1-8. [doi: 10.7790/ejap.v5i2.153]
43. Askjer S, Mathiasen K. The working alliance in blended versus face-to-face cognitive therapy for depression: a secondary analysis of a randomized controlled trial. Internet Interv 2021 Sep;25:100404 [FREE Full text] [doi: 10.1016/j.invent.2021.100404] [Medline: 34401364]
44. Mathiasen K, Andersen TE, Lichtenstein MB, Ehlers LH, Riper H, Kleiboer A, et al. The clinical effectiveness of blended cognitive behavioral therapy compared with face-to-face cognitive behavioral therapy for adult depression: randomized controlled noninferiority trial. J Med Internet Res 2022 Sep 07;24(9):e36577 [FREE Full text] [doi: 10.2196/36577] [Medline: 36069798]

45. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: results of a pragmatic randomized controlled trial. J Affect Disord 2018 Dec;227:455-462. [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
46. Bisson JI, Ariti C, Cullen K, Kitchiner N, Lewis C, Roberts NP, et al. Guided, internet based, cognitive behavioural therapy for post-traumatic stress disorder: pragmatic, multicentre, randomised controlled non-inferiority trial (RAPID). BMJ 2022 Jun 16;377:e069405 [FREE Full text] [doi: 10.1136/bmj-2021-069405] [Medline: 35710124]
47. Cloitre M, Amspoker AB, Fletcher TL, Hogan JB, Jackson C, Jacobs A, et al. Comparing the ratio of therapist support to internet sessions in a blended therapy delivered to trauma-exposed veterans: quasi-experimental comparison study. JMIR Ment Health 2022 Apr 27;9(4):e33080 [FREE Full text] [doi: 10.2196/33080] [Medline: 35475777]
48. Duffy D, Enrique A, Connell S, Connolly C, Richards D. Internet-delivered cognitive behavior therapy as a prequel to face-to-face therapy for depression and anxiety: a naturalistic observation. Front Psychiatry 2019 Jan 9;10:902 [FREE Full text] [doi: 10.3389/fpsyt.2019.00902] [Medline: 31998149]
49. Etzelmueller A, Radkovsky A, Hannig W, Berking M, Ebert DD. Patient's experience with blended video- and internet based cognitive behavioural therapy service in routine care. Internet Interv 2018 Jun;12:165-175 [FREE Full text] [doi: 10.1016/j.invent.2018.01.003] [Medline: 30135780]
50. Kemmeren LL, van Schaik A, Smit JH, Ruwaard J, Rocha A, Henriques M, et al. Unraveling the black box: exploring usage patterns of a blended treatment for depression in a multicenter study. JMIR Ment Health 2019 Jul 25;6(7):e12707 [FREE Full text] [doi: 10.2196/12707] [Medline: 31344670]
51. Kenter R, Warmerdam L, Brouwer-Dudokdewit C, Cuijpers P, van Straten A. Guided online treatment in routine mental health care: an observational study on uptake, drop-out and effects. BMC Psychiatry 2013 Jan 31;13:43 [FREE Full text] [doi: 10.1186/1471-244X-13-43] [Medline: 23368894]
52. Kenter RM, van de Ven PM, Cuijpers P, Koole G, Niamat S, Gerrits RS, et al. Costs and effects of internet cognitive behavioral treatment blended with face-to-face treatment: results from a naturalistic study. Internet Interv 2015 Mar;2(1):77-83. [doi: 10.1016/j.invent.2015.01.001]
53. Kooistra L, Ruwaard J, Wiersma J, van Oppen P, Riper H. Working alliance in blended versus face-to-face cognitive behavioral treatment for patients with depression in specialized mental health care. J Clin Med 2020 Jan 27;9(2):347 [FREE Full text] [doi: 10.3390/jcm9020347] [Medline: 32012722]
54. Lungu A, Jun JJ, Azarmanesh O, Leykin Y, Chen CE. Blended care-cognitive behavioral therapy for depression and anxiety in real-world settings: pragmatic retrospective study. J Med Internet Res 2020 Jul 06;22(7):e18723 [FREE Full text] [doi: 10.2196/18723] [Medline: 32628120]
55. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015 May 26;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
56. Mol M, Dozeman E, Provoost S, van Schaik A, Riper H, Smit JH. Behind the scenes of online therapeutic feedback in blended therapy for depression: mixed-methods observational study. J Med Internet Res 2018 May 03;20(5):e174 [FREE Full text] [doi: 10.2196/jmir.9890] [Medline: 29724708]
57. Nakao S, Nakagawa A, Oguchi Y, Mitsuda D, Kato N, Nakagawa Y, et al. Web-based cognitive behavioral therapy blended with face-to-face sessions for major depression: randomized controlled trial. J Med Internet Res 2018 Sep 21;20(9):e10743 [FREE Full text] [doi: 10.2196/10743] [Medline: 30249583]
58. Romijn G, Batelaan N, Koning J, van Balkom A, de Leeuw A, Benning F, et al. Acceptability, effectiveness and cost-effectiveness of blended cognitive-behavioural therapy (bCBT) versus face-to-face CBT (ftfCBT) for anxiety disorders in specialised mental health care: a 15-week randomised controlled trial with 1-year follow-up. PLoS One 2021;16(11):e0259493 [FREE Full text] [doi: 10.1371/journal.pone.0259493] [Medline: 34767575]
59. Tarp K, Nielsen AS. Patient perspectives on videoconferencing-based treatment for alcohol use disorders. Alcohol Treat Q 2017 Jul 17;35(4):344-358. [doi: 10.1080/07347324.2017.1348785]
60. Thase ME, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani G, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Mar 01;175(3):242-250 [FREE Full text] [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
61. van de Wal M, Thewes B, Gielissen M, Speckens A, Prins J. Efficacy of blended cognitive behavior therapy for high fear of recurrence in breast, prostate, and colorectal cancer survivors: the SWORD study, a randomized controlled trial. J Clin Oncol 2017 Jul 01;35(19):2173-2183. [doi: 10.1200/JCO.2016.70.5301] [Medline: 28471726]
62. Vernmark K, Hesser H, Topooco N, Berger T, Riper H, Luuk L, et al. Working alliance as a predictor of change in depression during blended cognitive behaviour therapy. Cogn Behav Ther 2019 Jul 29;48(4):285-299 [FREE Full text] [doi: 10.1080/16506073.2018.1533577] [Medline: 30372653]
63. Witlox M, Garnefski N, Kraaij V, de Waal MW, Smit F, Bohlmeijer E, et al. Blended acceptance and commitment therapy versus face-to-face cognitive behavioral therapy for older adults with anxiety symptoms in primary care: pragmatic single-blind cluster randomized trial. J Med Internet Res 2021 Mar 26;23(3):e24366 [FREE Full text] [doi: 10.2196/24366] [Medline: 33769293]

64. Wu MS, Chen SY, Wickham RE, O'Neil-Hart S, Chen C, Lungu A. Outcomes of a blended care coaching program for clients presenting with moderate levels of anxiety and depression: pragmatic retrospective study. JMIR Ment Health 2021 Oct 21;8(10):e32100 [FREE Full text] [doi: 10.2196/32100] [Medline: 34673534]
65. Wu MS, Wickham RE, Chen SY, Chen C, Lungu A. Examining the impact of digital components across different phases of treatment in a blended care cognitive behavioral therapy intervention for depression and anxiety: pragmatic retrospective study. JMIR Form Res 2021 Dec 17;5(12):e33452 [FREE Full text] [doi: 10.2196/33452] [Medline: 34927591]
66. Barkham M, Connell J, Stiles WB, Miles JN, Margison F, Evans C, et al. Dose-effect relations and responsive regulation of treatment duration: the good enough level. J Consult Clin Psychol 2006;74(1):160-167. [doi: 10.1037/0022-006x.74.1.160]
67. Pauley D, Cuijpers P, Papola D, Miguel C, Karyotaki E. Two decades of digital interventions for anxiety disorders: a systematic review and meta-analysis of treatment effectiveness. Psychol Med 2023 Jan 28;53(2):567-579 [FREE Full text] [doi: 10.1017/S0033291721001999] [Medline: 34047264]
68. Titzler I, Saruhanjan K, Berking M, Riper H, Ebert DD. Barriers and facilitators for the implementation of blended psychotherapy for depression: a qualitative pilot study of therapists' perspective. Internet Interv 2018 Jun;12:150-164 [FREE Full text] [doi: 10.1016/j.invent.2018.01.002] [Medline: 30135779]
69. Doukani A, Free C, Araya R, Michelson D, Cerga-Pashoja A, Kakuma R. Practitioners' experience of the working alliance in a blended cognitive-behavioural therapy intervention for depression: qualitative study of barriers and facilitators. BJPsych Open 2022 Jul 25;8(4):e142 [FREE Full text] [doi: 10.1192/bjo.2022.546] [Medline: 35876079]
70. Schuster R, Pokorny R, Berger T, Topooco N, Laireiter AR. The advantages and disadvantages of online and blended therapy: survey study amongst licensed psychotherapists in Austria. J Med Internet Res 2018 Dec 18;20(12):e11007. [doi: 10.2196/11007] [Medline: 30563817]
71. Davies F, Harris M, Shepherd HL, Butow P, Beatty L, Kemp E, et al. Promise unfulfilled: implementing web-based psychological therapy in routine cancer care, a qualitative study of oncology health professionals' attitudes. Psychooncology 2022 Jul 22;31(7):1127-1135. [doi: 10.1002/pon.5900] [Medline: 35129251]
72. Luo C, Sanger N, Singhal N, Pattrick K, Shams I, Shahid H, et al. Addendum to a comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: a systematic review and meta-analysis. EClinicalMedicine 2020 Oct;27:100580. [doi: 10.1016/j.eclinm.2020.100580]


## Abbreviations

**BT:** blended therapy
**CBT:** cognitive behavioral therapy
**ES:** effect size
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RCT:** randomized controlled trial


*Edited by T de Azevedo Cardoso; submitted 05.06.23; peer-reviewed by S Rennick-Egglestone, G Waller; comments to author 08.10.23; revised version received 27.08.24; accepted 17.09.24; published 29.10.24.*

*Please cite as:*
*Ferrao Nunes-Zlotkowski K, Shepherd HL, Beatty L, Butow P, Shaw JM*
*Blended Psychological Therapy for the Treatment of Psychological Disorders in Adult Patients: Systematic Review and Meta-Analysis*
*Interact J Med Res 2024;13:e49660*
*URL: https://www.i-jmr.org/2024/1/e49660*
*doi:10.2196/49660*
*PMID:39470720*

Review

# Screening for Depression Using Natural Language Processing: Literature Review

Bazen Gashaw Teferra[1], BSc, MSc, PhD; Alice Rueda[1], PhD; Hilary Pang[1], MSc, MD; Richard Valenzano[2], PhD; Reza Samavi[3], PhD; Sridhar Krishnan[3], PhD; Venkat Bhat[1,4], MSc, MD

[1]Unity Health Toronto, St. Michael's Hospital, Interventional Psychiatry Program, Toronto, ON, Canada

[2]Toronto Metropolitan University, Department of Computer Science, Toronto, ON, Canada

[3]Toronto Metropolitan University, Department of Electrical, Computer, and Biomedical Engineering, Toronto, ON, Canada

[4]University of Toronto, Department of Psychiatry, Toronto, ON, Canada

**Corresponding Author:**
Venkat Bhat, MSc, MD
Unity Health Toronto
St. Michael's Hospital
Interventional Psychiatry Program
193 Yonge Street, 6-012
Toronto, ON, M5B 1M4
Canada
Phone: 1 4163604000 ext 76404
Email: venkat.bhat@utoronto.ca

## Abstract

**Background:** Depression is a prevalent global mental health disorder with substantial individual and societal impact. Natural language processing (NLP), a branch of artificial intelligence, offers the potential for improving depression screening by extracting meaningful information from textual data, but there are challenges and ethical considerations.

**Objective:** This literature review aims to explore existing NLP methods for detecting depression, discuss successes and limitations, address ethical concerns, and highlight potential biases.

**Methods:** A literature search was conducted using Semantic Scholar, PubMed, and Google Scholar to identify studies on depression screening using NLP. Keywords included "depression screening," "depression detection," and "natural language processing." Studies were included if they discussed the application of NLP techniques for depression screening or detection. Studies were screened and selected for relevance, with data extracted and synthesized to identify common themes and gaps in the literature.

**Results:** NLP techniques, including sentiment analysis, linguistic markers, and deep learning models, offer practical tools for depression screening. Supervised and unsupervised machine learning models and large language models like transformers have demonstrated high accuracy in a variety of application domains. However, ethical concerns related to privacy, bias, interpretability, and lack of regulations to protect individuals arise. Furthermore, cultural and multilingual perspectives highlight the need for culturally sensitive models.

**Conclusions:** NLP presents opportunities to enhance depression detection, but considerable challenges persist. Ethical concerns must be addressed, governance guidance is needed to mitigate risks, and cross-cultural perspectives must be integrated. Future directions include improving interpretability, personalization, and increased collaboration with domain experts, such as data scientists and machine learning engineers. NLP's potential to enhance mental health care remains promising, depending on overcoming obstacles and continuing innovation.

***(Interact J Med Res 2024;13:e55067)*** doi:10.2196/55067

**KEYWORDS**

depression; natural language processing; NLP; sentiment analysis; machine learning; deep learning; transformer-based models; large language models; cross-cultural; research domain criteria; RDoC

## Introduction

### Background

Depression is a prevalent mental health disorder affecting 280 million people worldwide and accounts for more than 47 million disability-adjusted life-years [1,2]. Characterized by symptoms such as persistent feelings of sadness, diminished interest, and impaired daily functioning, depression can severely impact an individual's quality of life [3] and is associated with suicide and premature mortality from comorbidities [4]. However, traditional methods of diagnosing and screening for depression primarily rely on subjective clinical assessments, which can be time-consuming and susceptible to the inherent biases of health care professionals [5].

In recent years, researchers have explored advanced technologies like natural language processing (NLP) to address the limitations of traditional methods in detecting and understanding depression [6,7]. NLP, a field of artificial intelligence (AI), enables machines to automatically analyze and extract valuable insights from textual data [8]. This technology shows immense potential in transforming how we identify and manage mental health disorders. By leveraging the vast amount of digital information generated daily, including social media posts, electronic health records, and web-based forums, NLP can assist in detecting subtle linguistic cues and patterns that may indicate depressive symptoms [9].

The integration of NLP in detecting depression offers multiple advantages. Not only does it promise improved accuracy and efficiency, but it also brings the advantage of scalability. By analyzing large amounts of textual data on a population level, NLP enables a comprehensive understanding of depression, potentially leading to early detection and intervention. In addition, NLP-based approaches might contribute to the reduction of the stigma surrounding mental health [10] by providing a more objective and nonjudgmental assessment of individuals' emotional well-being. This shows that NLP can help create a more supportive environment for those dealing with depression.

However, while NLP shows significant potential for mental health care, we must also recognize the challenges and ethical considerations that come with its implementation. Issues like privacy concerns, data security, and potential biases demand critical analysis. In this literature review, we aim to explore the current state of research on using NLP techniques for detecting depression. We will discuss the successes and limitations of this rapidly evolving technology, along with its future scenarios in improving mental health diagnosis and care. By shedding light on this rapidly evolving field, our goal is to foster informed discussions and encourage further advancements that will ultimately benefit individuals living with depression while promoting more effective mental health screening systems.

### This Review

This literature review aimed to provide a broad overview of the potential and challenges of using NLP for depression screening by extracting valuable information from textual data. While we discuss various NLP techniques and their applications, the focus is on presenting a comprehensive view of the field rather than delving into technical details. Our goal is to offer readers a high-level understanding of the opportunities and limitations presented by NLP in detecting depression while also highlighting ethical considerations and future directions. By providing this broad perspective, we hope to foster further exploration and innovation in applying NLP to enhance mental health support systems.

## Methods

### Literature Search Strategy

A comprehensive literature search was conducted using 3 web-based databases—Semantic Scholar, PubMed, and Google Scholar. These databases were chosen for their extensive coverage of research in the fields of computer science, health care, and AI. The search aimed to identify studies focusing on depression screening using NLP techniques.

The search strategy involved using a combination of relevant keywords and Boolean operators. The following search terms were used: "depression screening" OR "depression detection" AND "natural language processing" OR "NLP." This search query was tailored to each database to ensure compatibility with their specific search functions.

The inclusion criteria for selecting studies were broad, encompassing a range of study designs, including original research articles, review papers, and technical reports. Studies were included if they discussed the application of NLP techniques for depression screening or detection, addressed the successes and limitations of such approaches, or explored ethical considerations and potential biases. No restrictions were placed on the publication date to ensure a comprehensive overview of the historical development and current state of the field.

### Study Selection and Data Extraction

The initial search yielded many results. To manage the screening process efficiently, the titles and abstracts of the retrieved studies were imported into reference management software (Zotero). Duplicates were removed, and the remaining studies were screened.

During the initial screening, studies were assessed based on their relevance to the research topic. Studies that did not specifically address depression screening or detection using NLP were excluded. Studies focusing solely on other mental health disorders without a clear connection to depression were also excluded.

In the study selection process, 2 independent reviewers initially screened titles and abstracts against the predefined inclusion and exclusion criteria. Discrepancies between the reviewers were identified and resolved through a consensus meeting, where both reviewers discussed their decisions and clarified any misunderstandings related to the criteria. If consensus could not be reached, a third independent reviewer was consulted to provide an additional perspective and make the final decision. The full texts of the remaining studies were then reviewed in detail by the 2 independent reviewers, following the same process described for the title and abstract screening. Studies

were included in the final selection if they provided substantial contributions to the understanding and application of NLP in depression screening. This included discussions on NLP techniques, depression detection methods, classification models, datasets, ethical considerations, cross-cultural perspectives, or future directions in the field.

Data extraction was performed concurrently with the full-text review. Relevant information from each study was extracted and organized into a structured format. This included details, such as the study's main objectives, methodologies used, key findings, limitations, and potential future directions suggested by the authors. The extracted data were then synthesized and analyzed to identify common themes and gaps in the literature, forming the basis for the discussion section of this literature review.

This review aimed to provide an up-to-date overview of the field by following the literature search strategy. It highlights the potential and challenges of using NLP for depression screening, along with ethical considerations and future research directions.

### Task Definition and Scope

#### *Overview*

The primary task addressed in this literature review is the detection or screening of depression using NLP techniques. This task involves the automatic analysis of textual data, such as social media posts, electronic health records, or clinical interview transcripts, to identify indicators of depressive symptoms and provide a classification or assessment of an individual's mental health status.

The scope of this review encompasses various subtasks and aspects related to depression detection, including the following: (1) classification, (2) severity classification, (3) depressive symptoms identification, (4) risk assessment, and (5) personalized depression analysis.

#### *Classification*

This involves categorizing textual data into depressive or nondepressive states, often using supervised machine-learning algorithms. The goal is to accurately distinguish between individuals experiencing depression and those who are not.

#### *Severity Classification*

Beyond binary classification, some studies focus on assessing the severity of depressive symptoms. This involves categorizing depression into different levels or stages, such as mild, moderate, or severe, based on the linguistic cues present in the text.

#### *Depressive Symptoms Identification*

NLP techniques are used to identify specific depressive symptoms, such as negative emotion, persistent feelings of sadness, changes in cognitive processes, or expressions of hopelessness. This task helps in understanding the nuanced emotional and cognitive states associated with depression.

#### *Risk Assessment*

Some studies aim to go beyond detection and focus on assessing the risk of depression-related outcomes, such as suicide risk or the likelihood of developing major depressive disorder. This task involves analyzing linguistic cues that may indicate a higher risk for adverse events.

#### *Personalized Depression Analysis*

There is a growing interest in personalized depression analysis, where NLP techniques are used to tailor interventions and treatments to individuals. This involves identifying unique linguistic patterns and behaviors associated with specific subgroups of depressed individuals.

By outlining these tasks and scope, we provide a clear framework for the literature review and ensure that the discussion remains focused on the application of NLP in depression detection and related areas.

## Results

### Overview

The literature search strategy described in the Methods section returned a diverse range of studies focusing on various aspects of depression detection using NLP techniques. These studies spanned different methodologies, including original research, review articles, and technical reports. The key findings from these studies focusing on the NLP techniques are summarized in Table 1 below and presented in a structured format for clarity. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram is presented in Figure 1 and the PRISMA checklist is presented in Multimedia Appendix 1.

**Table 1.** Summary of natural language processing (NLP) techniques for depression detection.

| NLP techniques | Methodology | Relevance to depression detection | Limitations | Study | Sample size, n | Dataset | Method | Results |
|---|---|---|---|---|---|---|---|---|
| Sentiment analysis | Analyzes the emotional tone of the text | Identifies negative language | Limited understanding of depression | Rathner et al [11], 2017 | 220 participants | Recruited participants were asked to reflect on their past year | LIWC[a] based features | $R^2$ value of 0.104 |
| Sentiment analysis | Analyzes the emotional tone of the text | Identifies negative language | Limited understanding of depression | Prabhu et al [12], 2022 | 189 sessions | DAIC-WOZ[b] | Word2vec feeds them as input to long short-term memory | 82.3% accuracy |
| Linguistic markers | Identifies linguistic features related to depression | Captures cognitive distortions and identifies the use of certain types of words | May overlook contextual complexities | Islam et al [13], 2018 | 7145 comments | Facebook user comments | Decision tree classifier from feature obtained through LIWC | *F*-measure of 0.71 |
| Linguistic markers | Identifies linguistic features related to depression | Captures cognitive distortions and identifies the use of certain types of words | May overlook contextual complexities | De Choudhury et al [14], 2021 | 554 users | Twitter | LIWC for determining 22 specific linguistic styles | 72.4% accuracy |
| Word embedding | Creates vectorized word representations | Preserves semantic relationships | May miss nuanced semantics | Stankevich et al [15], 2018 | 887 users | CLEF[c] and eRisk 2017 | Word embeddings and support vector machine model | $F_1$-score of 63.4% |
| Word embedding | Creates vectorized word representations | Preserves semantic relationships | May miss nuanced semantics | Lopez-Otero et al [16], 2017 | 189 sessions | DAIC-WOZ | GLoVe[d] vector inputs | $F_1$-score of 73% |
| Word embedding | Creates vectorized word representations | Preserves semantic relationships | May miss nuanced semantics | Mallol-Ragolta et al [17], 2019 | 189 sessions | DAIC-WOZ | GloVe embeddings | Unweighted average recall of 0.66 |
| Word embedding | Creates vectorized word representations | Preserves semantic relationships | May miss nuanced semantics | Dinkel et al [18], 2020 | 189 sessions | DAIC-WOZ | Pretrained word embeddings (ELMo[e]) | $F_1$-score of 84% |
| Word embedding | Creates vectorized word representations | Preserves semantic relationships | May miss nuanced semantics | Rutowski et al [19], 2020 | 16,000 sessions | American English spontaneous speech | GloVe word embedding | AUC[f] of 0.8 |
| Convolutional neural networks and recurrent neural networks | Captures local and sequential information in language data | Models language patterns of depressed individuals | Complex architectures, data-hungry | Korti and Kanakaraddi [7], 2022 | —[g] | Twitter | Recurrent neural network with long short-term memory | 91% accuracy |
| Convolutional neural networks and recurrent neural networks | Captures local and sequential information in language data | Models language patterns of depressed individuals | Complex architectures, data-hungry | Tejaswini et al [20], 2022 | 13,000 posts | Reddit and Twitter | Fasttext with long short-term memory | 87% accuracy |

| NLP techniques | Methodology | Relevance to depression detection | Limitations | Study | Sample size, n | Dataset | Method | Results |
|---|---|---|---|---|---|---|---|---|
| Large language models | Captures complex linguistic nuances and context | Achieves high-level understanding | Computationally expensive and requires specific fine-tuning | Senn et al [21], 2022 | 189 sessions | DAIC-WOZ | Fine-tuning BERT[h] and its variants | $F_1$-score of 0.62 |
| Large language models | Captures complex linguistic nuances and context | Achieves high-level understanding | Computationally expensive and requires specific fine-tuning | Hayati et al [22], 2022 | 53 participants | Interview questions | Few-shot learning on GPT[i]-3 | $F_1$-score of 0.64 |
| Large language models | Captures complex linguistic nuances and context | Achieves high-level understanding | Computationally expensive and requires specific fine-tuning | Németh et al [23], 2022 | Approximately 80,000 posts | Data acquired through SentiOne | Fine-tuning DistilBERT | 73% precision |

[a]LIWC: Linguistic Inquiry and Word Count.

[b]DAIC-WOZ: Distress Analysis Interview Corpus–Wizard-of-Oz set.

[c]CLEF: Conference and Labs of the Evaluation Forum.

[d]GLoVe: global vectors for word representation.

[e]ELMo: embeddings from language models.

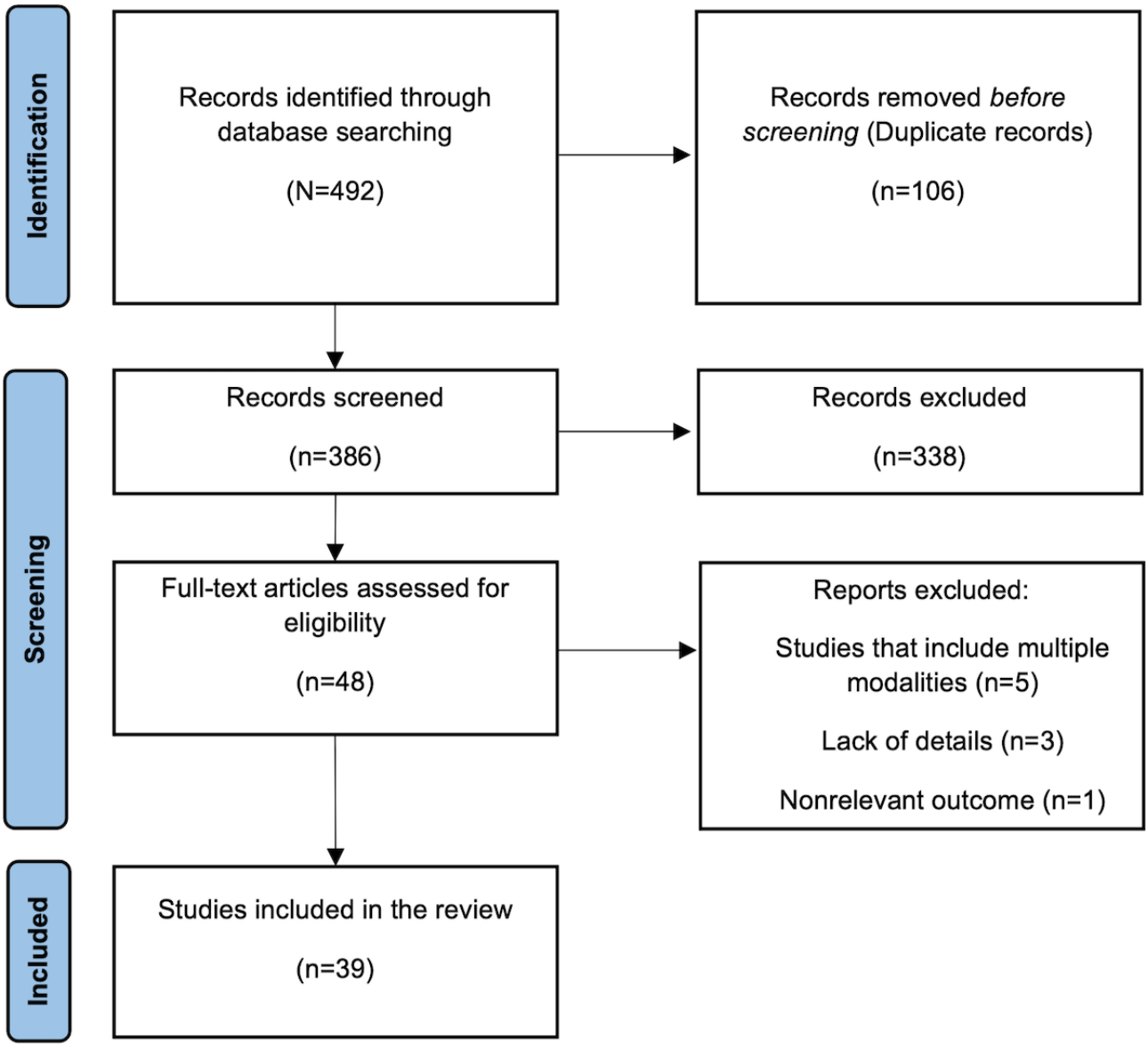
[f]AUC: area under the curve.

[g]Not available.

[h]BERT: bidirectional encoder representations from transformers.

[i]GPT: generative pretrained transformer.

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




## Historical Timeline of NLP Development and Relevance of Depression

NLP has gone through substantial development over the years, with technological advances and research contributing to its growth. Throughout its history, the field of NLP has continually expanded its capabilities, and the relevance of depression screening within this timeline has become increasingly noticeable.

### *Early Years and Rule-Based Systems (1950s-1970s)*

The origins of NLP can be traced back to the 1950s with the development of early computer programs like the Georgetown-IBM Experiment (developed jointly by Georgetown University and IBM) [24], which attempted to translate Russian sentences into English using basic rules and structures. In the 1970s, rule-based systems gained prominence. Systems like SHRDLU [25] demonstrated limited language understanding by manipulating blocks in a virtual world based on user commands. However, these systems had difficulty handling the complexity of natural language, including expressions of emotions and sentiment.

### *Knowledge-Based Approaches and Syntax Analysis (1980s-1990s)*

In the 1980s, there was a shift toward knowledge-based approaches, including expert systems [26]. Researchers attempted to encode linguistic rules and world knowledge to improve language understanding. In the 1990s, statistical methods, such as the hidden Markov model (HMM) [27] and part-of-speech tagging [28], gained popularity. These methods improved parsing and syntactic analysis, but the understanding of context, semantics, and emotions in language remained challenging.

### *Machine Learning and Statistical Methods (2000s-2020s)*

Machine learning algorithms, such as support vector machine and multi-layer perceptron improved performance on several NLP tasks, including sentiment analysis. Deep learning models, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), later enhanced NLP. Word embeddings

like Word2Vec and global vectors for word representation (GloVe) captured semantic relationships between words. Attention mechanisms and transformers, used in models like bidirectional encoder representations from transformers (BERT) and generative pretrained transformer (GPT), achieved remarkable results in language understanding and generation. These methods in relation to depression will be explained in detail in the Classification Models for Depression Detection: Machine Learning and Current State of the Art Models in Depression Detection sections.

### *Relevance of Depression in the NLP Timeline*

With the growth of web-based platforms and social media, textual data became abundant, and there was increased interest in the application of NLP for sentiment analysis and mood detection. Early efforts to identify emotional states and linguistic markers of depression emerged. The deep learning revolution then enabled more nuanced sentiment analysis and emotion recognition. Researchers started to explore the detection of mental health conditions, including depression, using NLP techniques. Studies focused on extracting linguistic cues related to depressive symptoms and emotional states from text data. Specific methods of depression detection using NLP will be further discussed in the following sections.

## NLP Techniques for Depression Detection

NLP techniques have been shown to be important in obtaining valuable insights from textual data acquired through social media, web-based forums, or textual health records for depression detection [13,20]. Given certain textual data, NLP can convert—through multiple steps—the textual data into a format that can point toward the presence or absence of depression. Among the initial steps in depression detection through NLP, *text preprocessing* and *feature extraction* play an essential role. Text preprocessing involves converting text data into a structured format suitable for analysis. Researchers have used techniques like tokenization, stemming, and lemmatization to achieve this [12]. Tokenization breaks down a certain transcript into individual words or tokens (which are parts of words). Stemming and lemmatization are both processes that involve reducing words to their base or root forms. Stemming often uses chopping (eg, jumps → jump, caring → car), while lemmatization applies language and context analysis for accurate reductions (eg, better → good, caring → care).

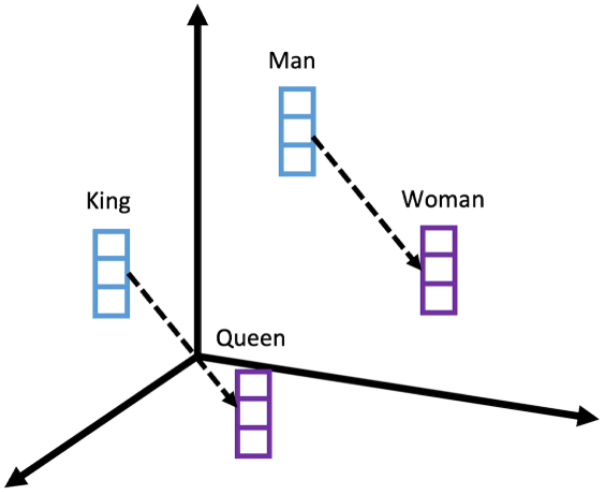
Another NLP technique that has been used to convert text into a more usable representation that has led to an increase in the accuracy of depression detection is feature extraction from text. Some examples of these methods include bag of words and term frequency-inverse document frequency [6]. The bag of words method creates a count-based representation of words present in a certain text by treating each word as a separate unit and, therefore, ignoring the order of the words. Term frequency-inverse document frequency assigns weights to words based on their count in a document and across the entire dataset, giving more importance to rare but distinctive words. These methods enable researchers to transform raw textual data into a quantitative representation, which can be used for further processing down the road.

One of the most widely used NLP techniques, and one that has been used as a proxy for identifying depression in language is *sentiment analysis* [29]. This approach examines the emotional tone of a text. Prior studies have shown a higher correlation between individuals with depression and the use of more negative words and the frequent expression of emotions related to sadness and hopelessness in their written language [30]. The linguistic inquiry and word count tool [31] is one example of such a sentiment analysis technique. This tool enables the automatic analysis of texts into preset word categories associated with depression—including negative emotions and cognitive processing [11]. By quantifying the emotional expressions in text data, researchers can gain valuable insights into the emotional state of individuals, which can in turn point toward the identification of potential depressive symptoms.

Researchers have also made substantial efforts to identify *linguistic markers and cues* that capture patterns related to depression. For example, prior studies have revealed that certain linguistic features, such as an increased use of first-person pronouns and a decreased use of third-person pronouns, can indicate the presence of depression [14,32]. In addition, the presence of cognitive distortions, characterized by negative thinking, has been found in the language of individuals with depression [33]. This suggests that by examining language patterns, NLP techniques offer a window into the cognitive and emotional processes underlying depression.

Recent advancements in NLP have enhanced the field of mental health disorder detection in general, especially using vectorized representations of language. *Word embeddings* and *contextual analysis using deep learning models* have substantially improved the accuracy and performance of depression detection models [34-37]. Word embeddings are created by transforming words into continuous vector representations, capturing semantic relationships and contextual meaning between words. To grasp the concept of word embeddings, envision a vector space where words are positioned based on their semantic meanings, allowing for intriguing relationships like “king” – “man” + “woman” = “queen” (see Figure 2 for the visualization of this analogy). Static word embeddings, such as GloVe [37], transform words into fixed vector representations that capture global semantic relationships, while dynamic embeddings, like embeddings from language model (ELMo) [38], provide context-dependent word representations. In the context of depression-related text data, previous studies have leveraged both GloVe [15-17] and ELMo [18] embeddings to capture word semantics and have achieved better accuracy in depression detection tasks. By preserving semantic relationships, word embeddings enable NLP models to better understand the meaning of words in context, which enhances the capacity to identify linguistic indicators of depression.

**Figure 2.** Visualization of how word embeddings capture analogy information from the words.




In summary, the use of NLP techniques, such as sentiment analysis, linguistic markers, and recent advancements like word embeddings contribute to a powerful toolkit for detecting depression from the text. These techniques have enabled researchers and clinicians to gain valuable insights into an individual's mental health state through their written and spoken languages, potentially giving rise to more accurate detection and intervention strategies for depression and other mental health disorders. As research in this domain continues to evolve, combining the strengths of classic NLP with cutting-edge developments—which will be described in the coming section—promises to enhance the understanding of depression further, leading to improved mental health outcomes for individuals worldwide.

## Classification Models for Depression Detection: Machine Learning

Machine learning models have emerged as a powerful tool in depression and mental health disorder detection in general, offering good capabilities to classify depressive and nondepressive states accurately. Leveraging the large amount of digital text data generated daily, these models have the potential to enhance mental health care by enabling more efficient and objective approaches to identifying and managing depression.

One approach is supervised machine learning where various algorithms have been applied to depression detection tasks, particularly binary classification (depressed or nondepressed) based on linguistic features extracted from text. Logistic regression, support vector machines, random forests, naive Bayes, and multi-layer perceptron classifiers are among the commonly used models, and they have shown promising results in accurately identifying and classifying depressive states [11,23,39]. These models use labeled datasets to learn patterns and relationships between textual features and depression status, contributing to more accurate and robust predictions.

Another set of approaches are the unsupervised machine learning techniques which have been used to uncover hidden structures within the depressed population. Techniques, such as K-means and hierarchical clustering, aim to identify distinct subgroups based on their linguistic patterns [40]. By grouping individuals with similar language use together, these clustering approaches have the potential to uncover different characteristics of depression. Such insights could lead to the development of personalized treatment strategies, catering to the unique needs of subgroups within the larger depressed population.

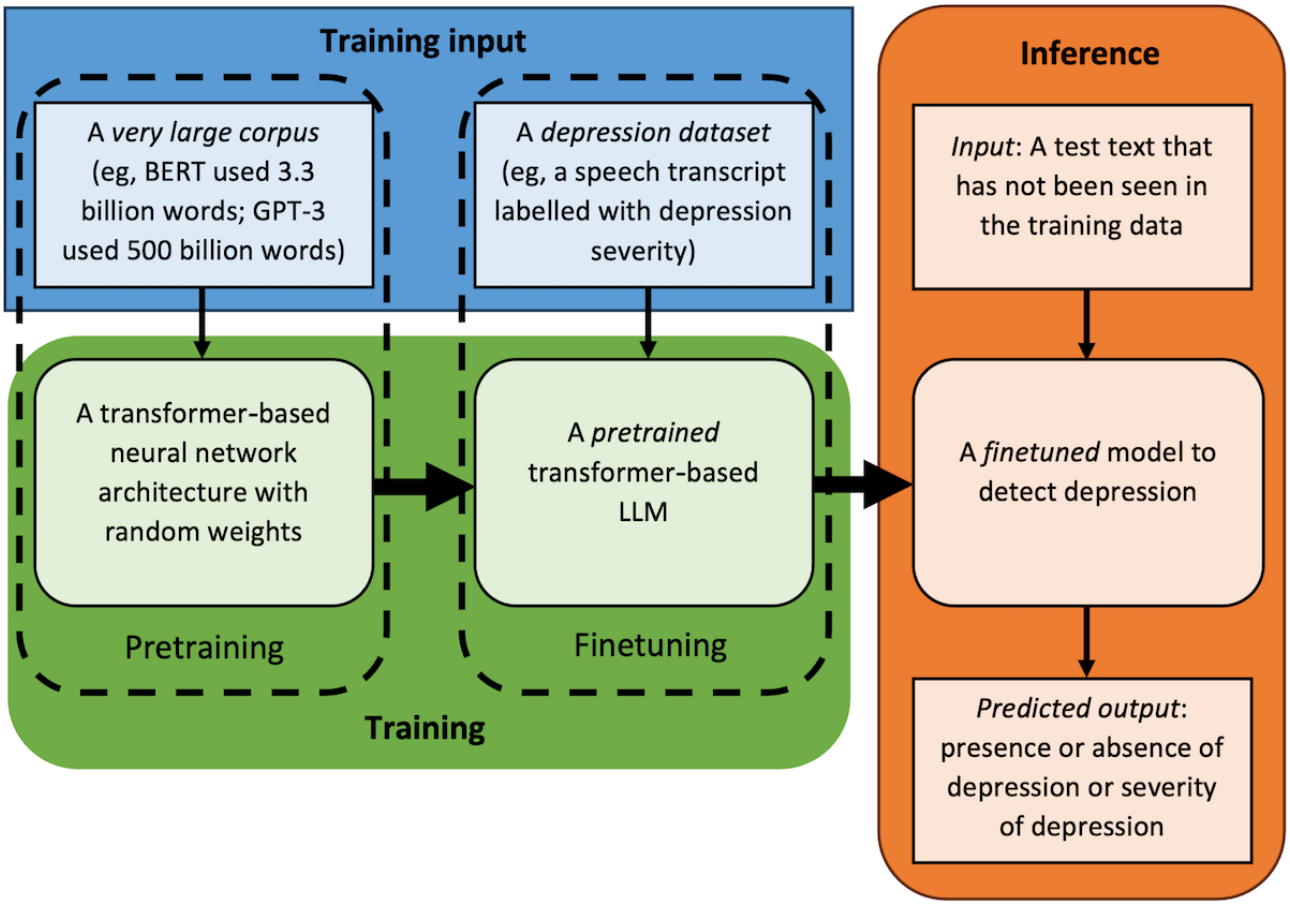
The introduction of deep learning has further advanced depression detection in the scope of NLP. For example, in the context of depression detection, CNNs and RNNs have gained popularity for their ability to capture sequential information and model temporal dependencies in language data [7,20]. CNNs effectively analyze local patterns within a text, while RNNs are well-suited for understanding the contextual dependencies that arise from sequential data. By capturing the linguistic cues related to depression, these deep learning architectures have substantially increased the performance of depression detection models from prior machine learning models. For example, Tejaswini et al [20] developed a novel approach called Fasttext convolution neural network with long short-term memory to detect depression in social media text data obtained from Reddit and Twitter. Their method achieved an 87% accuracy in distinguishing depression from nondepression in a dataset comprising of 13,000 samples, highlighting its potential for early detection of depressive states.

## Current State of the Art Models in Depression Detection

Transformer-based large language models (LLMs), another class of deep learning models, have also shown exceptional results in mental health detection tasks [41]. These models, characterized by their substantial number of parameters, often in the hundreds of millions to trillions, are artificial neural networks designed for natural language understanding and generation, using a self-attention mechanism to process input data in parallel and capture contextual information across long sequences. The transformer architecture [42] has been widely used in language modeling tasks and has proved to be highly effective in sentiment analysis [29,43-45]. In the context of depression detection, these models excel at understanding the complex and nuanced emotional language used by individuals experiencing depression [21,22,39].

One of the essential attributes of LLMs in the context of depression detection lies in their ability to do transfer learning—a technique that involves pretraining a model on a large and comprehensive dataset before fine-tuning it for a specific task. This paradigm has exhibited immense potential in the domain of depression detection [19]. Notable pretrained language models, such as BERT [46], MentalBERT [47], and GPT [48] have been fine-tuned on datasets curated from depression-related text sources. This strategic adaptation enables these models to grasp the details of the lexicon and linguistic context unique to depression [21,22]. The outcome is a refined model capable of discerning and contextualizing the language manifestations that accompany depression, thereby enhancing both accuracy and generalizability in depression detection models. Figure 3 shows the typical steps taken to generate a fine-tuned model from a pretrained model to make an inference and predict depression.

**Figure 3.** Steps to generate a typical large language model (LLM)–based depression detection model. BERT: bidirectional encoder representations from transformers; GPT: generative pretrained transformer.




By pretraining on vast and diverse linguistic datasets, these LLMs develop an innate comprehension of the structures and nuances of human language. This foundational knowledge enables them to decode the unique linguistic markers and emotional cues exhibited by individuals struggling with depression. The understanding of context enables these models to not only identify obvious expressions of depressive thoughts and emotions but also to identify the subtle and sometimes ambiguous linguistic signals that often evade traditional diagnostic approaches.

In summary, incorporating machine learning and deep learning models into depression detection has not only enhanced accuracy but also opened new avenues for understanding and managing mental health disorders. The ability of these models to identify patterns, relationships, and distinct subgroups within the depressed population offers valuable insights for early detection, personalized treatment, and more effective mental health interventions. As research in machine learning and deep learning continues to advance, the potential for further improving depression detection and mental health care becomes increasingly promising. Using the power of these technologies responsibly will be critical in realizing their full potential in enhancing mental health care and supporting individuals affected by depression. Table 1 summarizes the different NLP techniques discussed above.

## Datasets Used for Depression Detection

The availability of appropriate datasets plays a vital role in training and building NLP models, including for tasks, such as depression detection. In recent times, several publicly accessible datasets, which include a depression label, have emerged [49-55]. Some of the datasets used for depression detection are presented in Table 1.

One of the most popular datasets is the “distress analysis interview corpus/wizard-of-Oz set” [51]. This dataset consists of audio (which includes transcript) and video recordings that

simulate clinical interviews designed to assess distress levels related to depression, anxiety, and posttraumatic stress disorder. During these interviews, some participants act out distress, showcasing symptoms and experiences associated with depression, while others take on the role of interviewers, referred to as the "Wizard-of-Oz." Researchers use this dataset to develop and evaluate computational methods and algorithms for automatic distress detection and analysis, specifically focusing on cues related to depression-related distress. In another dataset, as presented in the study by Matcham et al [56], they presented data collected from 623 participants with a history of recurrent major depressive disorder. The study used smartphone sensors, wearable devices, and app-based questionnaires over 11 to 24 months. In addition, speech data were collected every 2 weeks through a speech task involving predetermined text and open-ended responses, with 82.2% of participants providing speech data.

However, despite these advancements, certain characteristics of the datasets present challenges when it comes to the trained models' ability to generalize. For example, small and imbalanced datasets (as observed in Jamil et al [55], where only 5% of the tweets contained a reference to depression) can cause underfitting or overfitting, such that the model becomes too tailored to the training data, making it less effective in handling new and unseen data [57]. In addition, the lack of diversity in the data, with most of the samples originating from specific demographics, may limit the applicability of the models across different populations. These limitations highlight the need for continuous improvement in dataset curation and selection to enhance the performance and applicability of depression detection models.

### Validation and Evaluation Metrics

Evaluating NLP models for depression detection involves carefully selecting appropriate metrics to assess their performance. In this context, some commonly used quantitative evaluation metrics include accuracy, precision, recall, $F_1$-score [58], and the area under the receiver operating characteristic curve [59]. The results of different studies using these metrics can be seen in Table 1. Researchers often use cross-validation and external validation strategies to ensure the models' generalizability and effectiveness.

However, being mindful of the potential limitations and biases associated with these quantitative metrics and validation techniques is essential. For instance, while accuracy is a commonly used metric, more informative evaluations metrices are needed for imbalanced datasets—where one class significantly outweighs the others. The model may achieve high accuracy in such cases by classifying most samples into the majority class [60], but might not reflect the actual performance of the prediction model.

Cross-validation techniques [61], which are widely used for model evaluation, also have limitations that need to be considered. One common approach involves splitting the dataset into 2 parts, using one for training the model and the other for testing its performance. However, this method may introduce higher bias, as crucial information from the unutilized data is left out during the training phase.

Another cross-validation technique, leave-one-out cross-validation, attempts to mitigate bias by training on the entire dataset while sequentially leaving out one data point for testing. Although this approach uses all data points, it can lead to higher variation in the testing phase, mainly if the omitted data point is an outlier. In addition, the leave-one-out cross-validation method demands substantial execution time as it iterates over the number of data points, making it computationally expensive for large datasets.

Alternatively, researchers can opt for k-fold cross-validation, where the dataset is divided into k subsets, and the model is trained on all except one subset reserved for evaluation. This approach helps reduce bias compared with simple two-part cross-validation, but it can still introduce variation in the testing phase due to the different subsets used for evaluation.

As researchers use cross-validation to assess NLP models for depression detection, they must be mindful of these limitations and biases. Ensuring accurate and reliable model assessment requires understanding the trade-offs of each method and carefully selecting the most appropriate validation technique based on the dataset's characteristics and research objectives. By being aware of these considerations, we can ensure more robust and trustworthy evaluations of NLP models in the crucial domain of depression detection.

Aside from quantitative metrics, qualitative evaluation metrics also play a substantial role in assessing the performance of NLP models for depression detection, as they offer insights beyond numerical measurements. While quantitative metrics provide valuable information about the models' accuracy and efficiency, qualitative evaluation metrics explore the models' interpretability, user experience, and overall impact.

One important qualitative evaluation metric is interpretability [62]. NLP models, especially those using complex deep learning techniques, are often considered "black boxes" because it is challenging to understand how they arrive at their predictions. However, interpretability is essential in applications related to mental health. Clinicians, researchers, and users must comprehend how the model reaches its conclusions to trust its decisions. Therefore, techniques that explain the model's predictions, such as attention mechanisms or feature visualization [63], are essential for ensuring the model's transparency and interpretability.

User experience and acceptability are also crucial qualitative metrics to consider. When deploying NLP-based depression detection systems in real-world settings, it is vital to measure end users' experiences—such as patients and therapists. Feedback from users can shed light on the system's usability, ease of integration into existing workflows, and its ability to provide valuable insights during the prediction. Exploring user perspectives can help improve the model's design, its practicality, and effectiveness in real-world mental health settings.

Ultimately, qualitative evaluation metrics complement quantitative assessments by offering a more comprehensive understanding of the NLP model's impact on mental health care. By considering interpretability and user experience, we

can develop more well-rounded and effective NLP-based depression detection systems that align with the needs and expectations of both patients and mental health professionals.

### Cultural and Multilingual Perspectives

The cultural and linguistic diversity surrounding mental health expressions presents distinctive challenges when it comes to detecting depression across different languages and cultures [64,65]. Researchers have recognized these challenges and sought to address them through cross-cultural research, highlighting the necessity for culturally sensitive depression detection models [66].

An illustrative example of such efforts is the study conducted by Lyu et al [67]. In this research, the focus was on depression detection using text-only social media data from the Chinese platform Weibo. The researchers considered a broader range of linguistic features that are relevant to depression, considering cultural factors and suicide risk specific to the Chinese language. To achieve this, they analyzed depression scores and past posts from 789 Weibo users. The outcome was a predictive model that showed promising results in detecting depression among the Chinese-speaking population.

This study served as an important reminder of the need for cultural expressions when it comes to improving the recognition of depression within specific linguistic and cultural groups. By considering the unique ways in which individuals from different languages and cultural backgrounds communicate their mental health experiences, we can improve the accuracy of depression detection models and make them more inclusive. As we continue to explore and expand our understanding of cross-cultural and multilingual perspectives on depression detection, we can move toward providing better mental health support and care for diverse populations around the world.

## Discussion

### Principal Findings

This literature review explored the potential of NLP in enhancing depression screening by analyzing textual data. The review revealed a range of NLP techniques, including sentiment analysis; linguistic markers; word embeddings; and deep learning models, such as CNNs, RNNs, and LLMs, that have been successfully applied to depression detection. The studies demonstrated the efficacy of these techniques, with machine learning models achieving high accuracy in classifying depressive states. However, ethical concerns, including privacy, bias, and interpretability, were also identified as critical challenges. In addition, the importance of cross-cultural and multilingual perspectives was emphasized, highlighting the need for culturally sensitive models. The review further discussed the integration of depression detection using NLP within the research domain criteria (RDoC) framework, mapping linguistic cues to psychological and biological constructs. Overall, the findings showcase the potential of NLP in enhancing mental health support systems while also presenting ethical and technical challenges that require continued innovation and collaboration.

### Comparison to Prior Work

The main findings of this review align with prior work in the field, which has also identified the potential of NLP techniques in mental health detection [9,10,29]. However, this review extends the understanding by incorporating the latest advancements in LLMs and their application to depression detection. Previous studies have largely focused on traditional machine learning techniques, while this review highlights the significant potential impact of deep learning and LLMs, as well as the comparison of current state-of-the-art models with prior classification models for depression detection.

### Integration of Depression Detection Using NLP Within the RDoC Framework

The RDoC framework provides a comprehensive approach to investigating mental health and psychopathology by focusing on fundamental psychological and biological systems rather than relying solely on traditional diagnostic categories. The RDoC framework acknowledges the complexity of mental health conditions and aims to foster new research approaches that lead to improved diagnosis, prevention, intervention, and treatment [68]. Here, we will explore how the detection of depression using NLP aligns with the principles and objectives of the RDoC framework.

The RDoC framework is organized into several major functional domains, each containing psychological and biological dimensions or constructs that span the range from normal to abnormal functioning [68]. Depression, a multifaceted mental health condition, touches upon multiple domains within the RDoC framework. These domains include negative valence systems, positive valence systems, cognitive systems, and social processes. NLP techniques for depression detection often analyze textual data to extract linguistic cues, sentiment, and cognitive patterns related to depression [14,29,30,32]. These patterns can be mapped onto constructs within the RDoC domains to gain insights into the underlying psychological and biological mechanisms associated with depressive symptoms.

For example, persistent negative emotions, such as sadness, hopelessness, and irritability, often characterize depression. This can be mapped to the negative valence systems domain in the RDoC framework. Some NLP methods that practice sentiment analysis for depression detection use linguistic markers associated with negative emotions and cognitive distortions [29,30]. Similarly, the cognitive systems domain in the RDoC framework focuses on cognitive processes and depression often involves cognitive distortions. NLP techniques can identify linguistic markers indicative of cognitive distortions within a text, connecting cognitive processes and linguistic expressions in depression [33].

Depression can also be mapped to the social processes domain within the RDoC framework which encompasses constructs related to social behavior, social cognition, and social communication. NLP methods offer insights into individuals' social expressions and interactions through their textual data, including social media posts and web-based forum discussions [20].

## Ethical Considerations and Limitations in NLP for Depression Detection

As NLP holds great potential in enhancing depression detection, it also gives rise to a range of ethical concerns that require careful consideration. Mental health data are highly sensitive. The information shared by individuals could include personal experiences, emotions, and medical history. Therefore, one of the main concerns includes safeguarding privacy and security of mental health data. To ensure the protection of individuals seeking support through NLP-based mental health services, it becomes crucial to use robust privacy and security measures. For privacy, even if data are anonymized, there may be a risk of reidentification. Clinical notes provide significant contextual information (eg, favorite movies and sports) that could be used by adversaries to be linked back to some other background information (eg, age group, general location information, etc) and identify the individual. Privacy regiments for NLP-based models, in addition to anonymization, should include guaranteed statistical privacy through differential privacy [69]. For security, in addition to cryptographic protection of data storage with support for provable worst-case security, an authentication and authorization regimen should be put in place to ensure the prevention of accidental or intentional unauthorized access.

Further ethical challenges arise from the potential biases that NLP models may inherit from the training data, giving rise to concerns regarding the robustness of depression detection [70-72]. Lack of robustness can have several adverse impacts when NLP is deployed in analyzing mental health text. A poor NLP model generalization (the distribution shift from the data the model is trained with to the data used in deployment) may lead to the model's failure to generalize well across various linguistic styles, terminologies, or expressions used by individuals. Lack of contextual understanding is another challenge as it might lead to the model losing essential contextual information, making it challenging to understand the full spectrum of an individual's mental state. Finally, bias and fairness issues are important aspects of model robustness. For instance, if certain demographic groups are underrepresented in the training data, it can lead to reduced accuracy for those specific populations, causing inequalities in health care.

In addition to these concerns, another substantial challenge arises from the integration of third-party application program interfaces and cloud-based services in NLP-based depression detection. While using third-party application program interfaces can enhance the capabilities of NLP models by, for example, incorporating pretrained language embeddings, it introduces a layer of dependency and potential security risks. These risks come from the need to share sensitive mental health data with external services, raising questions about data ownership, use, transparency, and compliance with privacy regulations.

A recent study by Straw et al [70] shed light on the integration of AI in health care. It underscored the critical need for collaboration between computer scientists and medical professionals to address biases in NLP models used in psychiatry. The research involved a comprehensive literature review of NLP applications in mental health, specifically evaluating biases in GloVe [37] and Word2Vec [36] word embeddings. The findings revealed significant biases related to religion, race, gender, nationality, sexuality, and age. Moreover, the review highlighted the need for more attention to these biases in existing research, signaling a limited cross-disciplinary collaboration in this domain.

Straw et al [70] emphasized the importance of addressing biases to prevent health gaps caused by AI and data-driven algorithms. They offered valuable recommendations for future research to minimize potential harm. By proactively working to identify and mitigate biases in NLP models, we can strive to create more equitable and just mental health support systems, ensuring that the benefits of NLP technology are accessible and effective for all individuals, regardless of their background or demographic characteristics. Textbox 1 summarizes some of the challenges and opportunities in using NLP for detecting depression.

**Textbox 1.** Challenges and opportunities in natural language processing for depression detection.

**Challenges and opportunities**

- Privacy concerns: robust anonymization techniques
- Biases: transparency and fairness
- Interpretability: explainable artificial intelligence
- User feedback: user-centric design
- Cross-cultural variations: culturally sensitive models
- Small, imbalanced datasets: data diversity

## Limitations of the Study

While this review provides valuable insights into the application of NLP for depression screening, some limitations should be acknowledged. First, the scope of the literature search was confined to 3 databases: Semantic Scholar, PubMed, and Google Scholar. Although these sources cover a wide range of academic publications, the exclusion of other major databases like IEEE Xplore and Scopus may have resulted in the omission of relevant studies, particularly those focusing on technical advancements in NLP. This may limit the breadth of the findings and leave certain innovations in NLP techniques underrepresented.

In addition, this review relies on qualitative synthesis without performing a meta-analysis of the selected studies. A quantitative meta-analysis could have provided a more robust statistical evaluation of the effectiveness of different NLP techniques in depression screening. Furthermore, while ethical

and cross-cultural issues were discussed, the review does not deeply analyze how these concerns are addressed across different studies. This omission could overlook critical gaps in ensuring the fairness and applicability of NLP models across diverse populations, limiting the generalizability of the findings.

### Future Directions

Despite the substantial potential that NLP has shown in depression detection, several challenges still lie ahead. The ambiguous characteristics of natural language and the ever-changing nature of language use in diverse contexts make it difficult to achieve consistently accurate detection. In addition, the risk of overfitting on small datasets and the necessity for large-scale, diverse datasets to ensure robust model training remain pressing concerns [41].

Future studies should focus on tackling these challenges to advance the field of NLP in depression detection. One important avenue for exploration is improving the interpretability of NLP models. As complex deep learning techniques become prevalent, it becomes increasingly crucial to understand and explain how these models arrive at their predictions. For example, the use of LLMs for interpretable mental health analysis has been explored in studies like that of Yang et al [73]. These models aim to provide not just predictions but also explainable insights into the mental health status of individuals. This direction holds the potential for improving the trustworthiness and clinical applicability of LLMs in depression screening. Future research should continue to address the challenges of accuracy, ethical considerations, and collaboration between NLP experts and mental health professionals to fully realize the benefits of LLMs in this context.

Personalization is another crucial aspect that future research should address. Depression is a complex and highly individualized condition, and so a one-size-fits-all approach may not be sufficient to meet the variable needs of individuals. Developing personalized depression detection tools that consider an individual's unique linguistic pattern, behavior, and context can enhance the accuracy of the detection process. These tools can cater to specific user requirements and provide tailored support and interventions.

Integrating user feedback and domain expertise in developing NLP models to achieve these advancements is crucial. Engaging with end users, including patients and mental health professionals, can provide valuable insights into the strengths and limitations of the models and help refine them to suit real-world applications better. Collaborating with domain experts, such as psychologists and psychiatrists, can ensure that the models align with clinical practices and address the most relevant aspects of depression detection and treatment.

Overall, the future of NLP in depression detection holds significant potential, but it also demands continued innovation and collaboration. By addressing challenges, improving interpretability, and personalizing depression detection tools, we can pave the way for more effective, user-centric, and clinically relevant solutions that contribute substantially to mental health care.

### Conclusions

In conclusion, adopting NLP techniques for depression detection holds significant potential in enhancing mental health support systems. This literature review has explored various NLP methodologies, applications, and challenges in detecting depression using textual data. While obstacles remain, ongoing advancements in NLP, ethical considerations, and cross-cultural insights pave the way for more accurate, accessible, and equitable mental health solutions.

The evolving field of NLP offers the potential for more effective detection of depression, but challenges persist. Overcoming the complexity of natural language and obtaining diverse datasets are key focus areas. Ethical considerations underscore the need for data privacy and model transparency. Integrating cross-cultural insights ensures culturally sensitive solutions that cater to diverse populations. With continued progress and collaboration, NLP can improve mental health care and well-being worldwide.

### Authors' Contributions

BGT created the outline, researched papers, and led the drafting of the initial manuscript, incorporating feedback from coauthors. AR provided detailed feedback on the draft. HP, RV, RS, SK, and VB critically reviewed the paper and offered valuable insights. VB supervised the study and provided guidance throughout the research process. The collaborative efforts of all authors have resulted in the final manuscript. All authors read and approved the final manuscript.

### Conflicts of Interest

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 & Behavior 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 Health care, American Foundation for Suicide Prevention, Roche Canada, Novartis, and Eisai. All other authors declare that they have no conflicts of interest.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 30 KB - ijmr_v13i1e55067_app1.docx ]

## References


1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
2. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2023-07-21]
3. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
4. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry 2015 Apr;72(4):334-341 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.2502] [Medline: 25671328]
5. Brown TA, Di Nardo PA, Lehman CL, Campbell LA. Reliability of DSM-IV anxiety and mood disorders: implications for the classification of emotional disorders. J Abnorm Psychol 2001;110(1):49-58. [doi: 10.1037/0021-843X.110.1.49]
6. Mali A, Sedamkar RR. Prediction of depression using machine learning and NLP approach. In: Proceedings of the International Conference on Intelligent Computing and Networking. 2021 Presented at: IC-ICN 2021; February 26-27, 2021; Mumbai, India. [doi: 10.1007/978-981-16-4863-2_15]
7. Korti SS, Kanakaraddi SG. Depression detection from Twitter posts using NLP and machine learning techniques. In: Proceedings of the Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology. 2022 Presented at: ICERECT 2022; December 26-27, 2022; Mandya, India. [doi: 10.1109/icerect56837.2022.10059773]
8. Chowdhary KR. Natural language processing. In: Fundamentals of Artificial Intelligence. New Delhi, India: Springer; Apr 05, 2020.
9. Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med 2022 Apr 08;5:46. [doi: 10.1038/s41746-022-00589-7]
10. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021 May 04;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
11. Rathner EM, Djamali J, Terhorst Y, Schuller B, Cummins N, Salamon G, et al. How did you like 2017? Detection of language markers of depression and narcissism in personal narratives. In: Proceedings of the Interspeech 2018. 2018 Presented at: Interspeech 2018; September 2-6, 2018; Hyderabad, India. [doi: 10.21437/interspeech.2018-2040]
12. Prabhu S, Mittal H, Varagani R, Jha S, Singh S. Harnessing emotions for depression detection. Pattern Anal Appl 2021 Sep 09;25:537-547. [doi: 10.1007/s10044-021-01020-9]
13. Islam MR, Kabir MA, Ahmed A, Kamal AR, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. Health Inf Sci Syst 2018 Aug 27;6(1):8 [FREE Full text] [doi: 10.1007/s13755-018-0046-0] [Medline: 30186594]
14. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. Proc Int AAAI Conf Web Soc Media 2021 Aug 03;7(1):128-137. [doi: 10.1609/icwsm.v7i1.14432]
15. Stankevich M, Isakov V, Devyatkin D, Smirnov I. Feature engineering for depression detection in social media. In: Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods. 2018 Presented at: ICPRAM 2018; January 16-18, 2018; Madeira, Portugal. [doi: 10.5220/0006598604260431]
16. Lopez-Otero P, Docio-Fernandez L, Abad A, Garcia-Mateo C. Depression detection using automatic transcriptions of de-identified speech. In: Proceedings of the Interspeech 2017. 2017 Presented at: Interspeech 2017; August 20-24, 2017; Stockholm, Sweden. [doi: 10.21437/interspeech.2017-1201]
17. Mallol-Ragolta A, Zhao Z, Stappen L, Cummins N, Schuller B. A hierarchical attention network-based approach for depression detection from transcribed clinical interviews. In: Proceedings of the Interspeech 2019. 2019 Presented at: Interspeech 2019; September 15-19, 2019; Graz, Austria. [doi: 10.21437/interspeech.2019-2036]
18. Dinkel H, Wu M, Yu K. Text-based depression detection on sparse data. arXiv. Preprint posted online on April 8, 2019 [FREE Full text] [doi: 10.48550/arXiv.1904.05154]
19. Rutowski T, Shriberg E, Harati A, Lu Y, Chlebek P, Oliveira R. Depression and anxiety prediction using deep language models and transfer learning. In: Proceedings of the 7th International Conference on Behavioural and Social Computing. 2020 Presented at: BESC 2020; November 5-7, 2020; Bournemouth, UK. [doi: 10.1109/besc51023.2020.9348290]
20. Tejaswini V, Sathya Babu K, Sahoo B. Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model. ACM Trans Asian Low Resour Lang Inf Process 2024 Jan 15;23(1):1-20. [doi: 10.1145/3569580]
21. Senn S, Tlachac ML, Flores R, Rundensteiner E. Ensembles of BERT for depression classification. In: Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2022 Presented at: EMBC 2022; July 11-15, 2022; Scotland, UK. [doi: 10.1109/embc48229.2022.9871120]
22. Hayati MF, Md. Ali MA, Md. Rosli AN. Depression detection on Malay dialects using GPT-3. In: Proceedings of the IEEE-EMBS Conference on Biomedical Engineering and Sciences. 2022 Presented at: IECBES 2022; December 7-9, 2022; Kuala Lumpur, Malaysia. [doi: 10.1109/iecbes54088.2022.10079554]

23. Németh R, Máté F, Katona E, Rakovics M, Sik D. Bio, psycho, or social: supervised machine learning to classify discursive framing of depression in online health communities. Qual Quant 2022 Jan 08;56:3933-3955. [doi: 10.1007/s11135-021-01299-0]
24. Hutchins WJ. The Georgetown-IBM experiment demonstrated in January 1954. In: Proceedings of the 6th Conference of the Association for Machine Translation in the Americas. 2004 Presented at: AMTA 2004; September 28-October 2, 2004; Washington, DC. [doi: 10.1007/978-3-540-30194-3_12]
25. Winograd T. Procedures as a representation for data in a computer program for understanding natural language. Massachusetts Institute of Technology Libraries. 1971. URL: https://dspace.mit.edu/handle/1721.1/7095 [accessed 2023-08-15]
26. Bench-Capon T. Expert systems. In: Knowledge Representation: An Approach to Artificial Intelligence. Cambridge, MA: Academic Press; 1990.
27. Elliott RJ, Moore JB, Aggoun L. Hidden Markov model processing. In: Hidden Markov Models: Estimation and Control. New York, NY: Springer; 1995.
28. Voutilainen A. Part-of-speech tagging. In: The Oxford Handbook of Computational Linguistics. Oxford, UK: Oxford Academic Press; Sep 18, 2012.
29. Babu NV, Kanaga EG. Sentiment analysis in social media data for depression detection using artificial intelligence: a review. SN Comput Sci 2022;3(1):74 [FREE Full text] [doi: 10.1007/s42979-021-00958-1] [Medline: 34816124]
30. Al-Mosaiwi M, Johnstone T. In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clin Psychol Sci 2018 Jan 05;6(4):529-542. [doi: 10.1177/2167702617747074]
31. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. The University of Texas at Austin. 2015. URL: https://repositories.lib.utexas.edu/server/api/core/bitstreams/b0d26dcf-2391-4701-88d0-3cf50ebee697/content [accessed 2024-10-17]
32. Rude S, Gortner EM, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cognit Emot 2004;18(8):1121-1133. [doi: 10.1080/02699930441000030]
33. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Commun 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
34. Bengio S, Heigold G. Word embeddings for speech recognition. In: Proceedings of the Interspeech 2014. 2014 Presented at: Interspeech 2014; September 14-18, 2014; Singapore, Singapore. [doi: 10.21437/interspeech.2014-273]
35. Li Y, Yang T. Word embedding for understanding natural language: a survey. In: Srinivasan S, editor. Guide to Big Data Applications. Cham, Switzerland: Springer; 2018.
36. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. Preprint posted online on January 16, 2013 . [doi: 10.48550/arXiv.1301.3781]
37. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP 2014; October 25-29, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1162]
38. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep contextualized word representations. arXiv. Preprint posted online on February 15, 2018 [FREE Full text] [doi: 10.48550/arXiv.1802.05365]
39. Alsagri HS, Ykhlef M. Machine learning-based approach for depression detection in Twitter using content and activity features. IEICE Trans Inf Syst 2020;E103.D(8):1825-1832. [doi: 10.1587/transinf.2020EDP7023]
40. Pinto SJ, Parente M. Comprehensive review of depression detection techniques based on machine learning approach. Soft Comput 2024 Jul 23. [doi: 10.1007/s00500-024-09862-1]
41. Teferra BG, Rose J. Predicting generalized anxiety disorder from impromptu speech transcripts using context-aware transformer-based neural networks: model evaluation study. JMIR Ment Health 2023 Mar 28;10:e44325 [FREE Full text] [doi: 10.2196/44325] [Medline: 36976636]
42. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv. Preprint posted online on June 12, 2017 . [doi: 10.48550/arXiv.1706.03762]
43. Liu KL, Li WJ, Guo M. Emoticon smoothed language models for Twitter sentiment analysis. Proc AAAI Conf Artif Intell 2021 Sep 20;26(1):1678-1684. [doi: 10.1609/aaai.v26i1.8353]
44. Xu H, Liu B, Shu L, Yu PS. DomBERT: domain-oriented language model for aspect-based sentiment analysis. arXiv. Preprint posted online on April 28, 2020 . [doi: 10.48550/arXiv.2004.13816]
45. Barbieri F, Anke LE, Camacho-Collados J. XLM-T: multilingual language models in Twitter for sentiment analysis and beyond. arXiv. Preprint posted online on April 25, 2021 [FREE Full text] [doi: 10.48550/arXiv.2104.12250]
46. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online on October 11, 2018 . [doi: 10.48550/arXiv.1810.04805]
47. Ji S, Zhang T, Ansari L, Fu J, Tiwari P, Cambria E. MentalBERT: publicly available pretrained language models for mental healthcare. arXiv. Preprint posted online on October 29, 2021 [FREE Full text] [doi: 10.48550/arXiv.2110.15621]
48. Floridi L, Chiriatti M. GPT-3: its nature, scope, limits, and consequences. Minds Mach 2020 Nov 01;30:681-694. [doi: 10.1007/s11023-020-09548-1]
49. Uddin MZ, Dysthe KK, Følstad A, Brandtzaeg PB. Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Comput Appl 2021 Aug 27;34:721-744. [doi: 10.1007/s00521-021-06426-4]

50. Shen Y, Yang H, Lin L. Automatic depression detection: an emotional audio-textual corpus and a Gru/Bilstm-based model. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. 2022 Presented at: ICASSP 2022; May 23-27, 2022; Singapore, Singapore. [doi: 10.1109/icassp43922.2022.9746569]
51. Valstar M, Gratch J, Schuller B, Ringeval F, Lalanne D, Torres M, et al. AVEC 2016: depression, mood, and emotion recognition workshop and challenge. In: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. 2016 Presented at: AVEC '16; October 16, 2016; Amsterdam, The Netherlands. [doi: 10.1145/2988257.2988258]
52. Ríssola EA, Bahrainian SA, Crestani F. A dataset for research on depression in social media. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. 2020 Presented at: UMAP '20; July 14-17, 2020; Genoa, Italy. [doi: 10.1145/3340631.3394879]
53. Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M. CLPsych 2015 shared task: depression and PTSD on Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: CLPsych@NAACL-HLT 2015; June 5, 2015; Denver, CO. [doi: 10.3115/v1/w15-1204]
54. Losada DE, Crestani F, Parapar J. Early detection of risks on the internet: an exploratory campaign. In: Proceedings of the 41st European Conference on IR Research. 2019 Presented at: ECIR 2019; April 14-18, 2019; Cologne, Germany. [doi: 10.1007/978-3-030-15719-7_35]
55. Jamil Z, Inkpen D, Buddhitha P, White K. Monitoring tweets for depression to detect at-risk users. In: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality. 2017 Presented at: CLPsych@ACL 2017; August 3, 2017; Vancouver, Canada. [doi: 10.18653/v1/w17-3104]
56. Matcham F, Leightley D, Siddi S, Lamers F, White KM, Annas P, et al. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Feb 21;22(1):136 [FREE Full text] [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]
57. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. Proc Natl Acad Sci U S A 2019 Aug 06;116(32):15849-15854 [FREE Full text] [doi: 10.1073/pnas.1903070116] [Medline: 31341078]
58. Powers DM. Evaluation: from precision, recallF-measure to ROC, informedness, markedness and correlation. arXiv. Preprint posted online on October 11, 2020 [FREE Full text] [doi: 10.48550/arXiv.2010.16061]
59. Naidu G, Zuva T, Sibanda EM. A review of evaluation metrics in machine learning algorithms. In: Proceedings of 12th Computer Science On-line Conference 2023. 2023 Presented at: CSOC 2023; April 3-5, 2023; Online. [doi: 10.1007/978-3-031-35314-7_2]
60. Brownlee J. Failure of classification accuracy for imbalanced class distributions. Machine Learning Mastery. 2021 Jan 22. URL: https://machinelearningmastery.com/failure-of-accuracy-for-imbalanced-class-distributions/ [accessed 2023-07-26]
61. Refaeilzadeh P, Tang L, Liu H. Cross-validation. In: Liu L, Özsu M, editors. Encyclopedia of Database Systems. New York, NY: Springer; Dec 14, 2016.
62. Shaban-Nejad A, Michalowski M, Buckeridge DL. Explainability and interpretability: keys to deep medicine. In: Explainable AI in Healthcare and Medicine. Cham, Switzerland: Springer; Nov 03, 2020.
63. cdpierse / transformers-interpret. GitHub. URL: https://github.com/cdpierse/transformers-interpret [accessed 2024-10-17]
64. Furler J, Kokanovic R. Mental health - cultural competence. Aust Fam Physician 2010 Apr;39(4):206-208. [Medline: 20372678]
65. Bredström A. Culture and context in mental health diagnosing: scrutinizing the DSM-5 revision. J Med Humanit 2019 Sep;40(3):347-363 [FREE Full text] [doi: 10.1007/s10912-017-9501-1] [Medline: 29282590]
66. Kalibatseva Z, Leong FT. A critical review of culturally sensitive treatments for depression: recommendations for intervention and research. Psychol Serv 2014 Nov;11(4):433-450. [doi: 10.1037/a0036047] [Medline: 25383996]
67. Lyu S, Ren X, Du Y, Zhao N. Detecting depression of Chinese microblog users via text analysis: combining linguistic inquiry word count (LIWC) with culture and suicide related lexicons. Front Psychiatry 2023 Feb 9;14:1121583 [FREE Full text] [doi: 10.3389/fpsyt.2023.1121583] [Medline: 36846219]
68. Research domain criteria (RDoC). National Institute of Mental Health. URL: https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc [accessed 2023-08-16]
69. Dwork C. Differential privacy. In: Proceedings of the 33rd International Colloquium on Automata, Languages and Programming. 2006 Presented at: ICALP 2006; July 10-14, 2006; Venice, Italy. [doi: 10.1007/11787006]
70. Straw I, Callison-Burch C. Artificial intelligence in mental health and the biases of language based models. PLoS One 2020 Dec 17;15(12):e0240376 [FREE Full text] [doi: 10.1371/journal.pone.0240376] [Medline: 33332380]
71. Caliskan A. Detecting and mitigating bias in natural language processing. Brookings. 2021 May 10. URL: https://www.brookings.edu/articles/detecting-and-mitigating-bias-in-natural-language-processing/ [accessed 2023-07-24]
72. Is bias in NLP models an ethical problem? Edlitera. 2023 Mar 29. URL: https://www.edlitera.com/blog/posts/bias-in-nlp [accessed 2023-07-24]
73. Yang K, Ji S, Zhang T, Xie Q, Kuang Z, Ananiadou S. Towards interpretable mental health analysis with large language models. arXiv. Preprint posted online on April 6, 2023 [FREE Full text] [doi: 10.48550/arXiv.2304.03347]

## Abbreviations

**AI:** artificial intelligence
**BERT:** bidirectional encoder representations from transformers
**CNN:** convolutional neural network
**ELMo:** embeddings from language model
**GloVe:** global vectors for word representation
**GPT:** generative pretrained transformer
**HMM:** hidden Markov model
**LLM:** large language model
**NLP:** natural language processing
**RDoC:** research domain criteria
**RNN:** recurrent neural network

*Edited by T de Azevedo Cardoso; submitted 01.12.23; peer-reviewed by T Zhang, M Rizvi, C Haag; comments to author 23.05.24; revised version received 25.06.24; accepted 17.09.24; published 04.11.24.*

*Please cite as:*
*Teferra BG, Rueda A, Pang H, Valenzano R, Samavi R, Krishnan S, Bhat V*
*Screening for Depression Using Natural Language Processing: Literature Review*
*Interact J Med Res 2024;13:e55067*
*URL: https://www.i-jmr.org/2024/1/e55067*
*doi:10.2196/55067*
*PMID:*

Review

# Methods Used in Co-Creation Within the Health CASCADE Co-Creation Database and Gray Literature: Systematic Methods Overview

Danielle Marie Agnello[1], BS, MSc; George Balaskas[2,3], BSc, MSc; Artur Steiner[4], BA, MA, PhD; Sebastien Chastin[1,5], BSc, MSc, PhD

[1]School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, United Kingdom

[2]Institute of Informatics and Telecommunications, National Centre of Scientific Research Demokritos, Athens, Greece

[3]Department of Digital Systems, University of Piraeus, Piraeus, Greece

[4]Yunus Centre for Social Business and Health, Glasgow Caledonian University, Glasgow, United Kingdom

[5]Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

**Corresponding Author:**
Danielle Marie Agnello, BS, MSc
School of Health and Life Sciences
Glasgow Caledonian University
Cowcaddens Rd
Glasgow, G4 0BA
United Kingdom
Phone: 44 7871788785
Email: danielle.agnello@gcu.ac.uk

## Abstract

**Background:** Co-creation is increasingly recognized for its potential to generate innovative solutions, particularly in addressing complex and wicked problems in public health. Despite this growing recognition, there are no standards or recommendations for method use in co-creation, leading to confusion and inconsistency. While some studies have examined specific methods, a comprehensive overview is lacking, limiting the collective understanding and ability to make informed decisions about the most appropriate methods for different contexts and research objectives.

**Objective:** This study aimed to systematically compile and analyze methods used in co-creation to enhance transparency and deepen understanding of how co-creation is practiced.

**Methods:** To enhance transparency and deepen understanding of how co-creation is practiced, this study systematically inventoried and analyzed methods used in co-creation. We conducted a systematic methods overview, applying 2 parallel processes: one within the peer-reviewed Health CASCADE Co-Creation Database and another within gray literature. An artificial intelligence–assisted recursive search strategy, coupled with a 2-step screening process, ensured that we captured relevant methods. We then extracted method names and conducted textual, comparative, and bibliometric analyses to assess the content, relationship between methods, fields of research, and the methodological underpinnings of the included sources.

**Results:** We examined a total of 2627 academic papers and gray literature sources, with the literature primarily drawn from health sciences, medical research, and health services research. The dominant methodologies identified were co-creation, co-design, coproduction, participatory research methodologies, and public and patient involvement. From these sources, we extracted and analyzed 956 co-creation methods, noting that only 10% (n=97) of the methods overlap between academic and gray literature. Notably, 91.3% (230/252) of the methods in academic literature co-occurred, often involving combinations of multiple qualitative methods. The most frequently used methods in academic literature included surveys, focus groups, photo voice, and group discussion, whereas gray literature highlighted methods such as world café, focus groups, role-playing, and persona.

**Conclusions:** This study presents the first systematic overview of co-creation methods, providing a clear understanding of the diverse methods currently in use. Our findings reveal a significant methodological gap between researchers and practitioners, offering insights into the relative prevalence and combinations of methods. By shedding light on these methods, this study helps bridge the gap and supports researchers in making informed decisions about which methods to apply in their work. Additionally, it offers a foundation for further investigation into method use in co-creation. This systematic investigation is a valuable resource for anyone engaging in co-creation or similar participatory methodologies, helping to navigate the diverse landscape of methods.

*(Interact J Med Res 2024;13:e59772)* doi:10.2196/59772

**KEYWORDS**

co-creation; coproduction; co-design; methods; participatory; inventory; text mining; methodology; research methods; CASCADE; research methods

# Introduction

## Overview

Co-creation has emerged as a powerful approach for fostering collaboration and innovation across various disciplines [1]. Co-creation has the potential to produce new or improved tailored practices and solutions, which address complex challenges and generate meaningful outcomes for a defined need [2,3]. In health care, co-creation is thought of as a potential way to bridge the bench-to-bedside divide [4,5] by developing interventions that are more acceptable and contextually appropriate, thereby potentially enhancing their effectiveness and impact on health outcomes [6]. It can be used to address complex problems in public health (eg, the obesity epidemic, persisting poverty, or food insecurity), which are particularly resistant to resolution [1,4]. Additionally, co-creation has been widely used in health care at a rapidly expanding rate [7].

At the core of co-creation is the active engagement of multiple stakeholders in a collective intelligence process to collaboratively design and implement projects, processes, services, and solutions [7-9]. Rooted in the potential for tapping into diverse perspectives, skills, expertise, and experiences, co-creation stands as a dynamic approach to fostering innovation and stakeholder engagement [10] and an effective means to avoid top-down approaches or mere stakeholder consultations [5].

There is a growing recognition of co-creation as a valuable methodology, yet, to date, there is no common practice or standardized approach [1,11]. Co-creation has increasing traction in terms of involving stakeholders, but the extent to which this creates a fundamental change in practices is still unclear, and there is a strong potential that it will remain tokenistic [12]. Major challenges remain and need to be confronted to make co-creation trustworthy and unlock its full potential [1,2,11-13].

## Fragmentation of Co-Creation

A well-documented challenge in co-creation research is its fragmentation. Slattery et al [14] describe the literature as complex, contradictory, and poorly synthesized, noting that inconsistent terminology makes it difficult to retrieve and understand relevant studies. Agnello et al [1] further highlight this fragmentation, attributing it to inconsistencies in terminology and inadequate cross-referencing and knowledge sharing among co-creation researchers. Smith et al [15] found that researchers operationalize co-creation in various ways, while Grindell et al [16] emphasize the interchangeable and often ambiguous use of the co-creation, co-design, and coproduction terms. These issues hinder the identification of best practices and the establishment of robust evidence for co-creation.

## Methods for Co-Creation

Research methods provide a structured and systematic approach to gathering and analyzing data, ensuring that findings are valid, reliable, and generalizable. Detailed insight into these processes is essential to truly understand and assess the quality of research [14,15,17-19]. Furthermore, well-documented research methods enable others to replicate studies and build on findings, which is crucial for advancing knowledge in any field [17]. Using appropriate methods allows researchers to control potential biases that might affect their results, making their findings more trustworthy and credible [20].

However, co-creation has been plagued with poor reporting of the steps involved in the process, in particular the methods used. Smith et al [15] urged the research community to improve reporting by clearly and consistently documenting methods. Similarly, Slattery et al [14] highlighted the need for better reporting on methods used in co-creation as a foundation for understanding its effectiveness and cost. An et al [13] also discovered limited reporting of methods in the co-creation literature and Lee et al [21] drew attention to the difficulty in evaluating co-creation due to a lack of systematic comprehension of methods. This challenge is also documented by Durugbo and Pawar [22], who highlighted the absence of a detailed description of the methods used to facilitate the interaction between the convener and the cocreators.

More than any other aspects of research, methods have the potential for adaptation or reuse across different contexts, research questions, and disciplines [6,17,23]. Consequently, poor reporting or restricted access to reliable methods creates inefficiencies and can slow progress. Given the numerous benefits of co-creation, it is surprising that there are no clear recommendations regarding methods appropriate for ensuring accurate, impactful, and trustworthy co-creation. The transferability of these methods suggests they could adapt the co-creation process to various contexts and research questions. However, poor reporting and limited access to trustworthy evidence about co-creation methods pose significant barriers, making it difficult to contextualize research, evaluate co-creation, replicate studies, or scale up co-creation research [1,14,15]. This lack of clarity also limits researchers' ability to build on existing knowledge and apply best practices [24].

While some studies have highlighted individual methods for co-creation, a comprehensive inventory or systematic overview of the full range of methods used in co-creation is lacking. Researchers striving to apply co-creation often lack the time to systematically search, retrieve, review, and compare all the available literature to develop a thorough and critical sense of the varied methods [1,25]. Therefore, a systematic methods overview could increase clarity and enhance collective understanding of which methods have been used in co-creation [25,26].

As Slattery et al [14] noted, there is considerable value in a greater synthesis and differentiation of the co-creation literature, highlighting that methodological and theoretical barriers in co-creation often prevent systematic comparison using traditional means. Consequently, this study aims to address this gap using a nontraditional means, by systematically inventorying and analyzing methods used in co-creation. By providing a comprehensive overview and analysis, this research will lay the groundwork for developing a robust taxonomy and compendium of co-creation methods in future research.

# Methods

## Definitions

We expanded on the definition of participatory methods by Vaughn et al [27] to establish a definition of methods used in co-creation: "Co-creation methods encompass a diverse range of tools, activities, approaches, and techniques strategically employed across the entirety of the co-creation process. These methods serve various purposes, including but not limited to data collection, facilitation, recruitment, reflection, data analysis, and dissemination, allowing for flexibility in achieving diverse objectives."

We defined co-creation according to Agnello et al [1]: "Co-creation is any act of collective creativity that involves a broad range of relevant and affected actors in creative problem-solving that aims to produce a desired outcome."

## Systematic Methods Overview

### *Overview*

Reviews on methods topics, known as methods overviews, are valuable for advancing research methods, making a systematic methods overview the ideal approach for thoroughly investigating methods used in co-creation [25,26,28]. This approach aims to synthesize guidance on methods from the literature; therefore, principles and strategies for a systematic approach were sourced from Gentles et al [25]. The strategies sourced from Gentles et al [25] and how they were applied in this study are described in Multimedia Appendix 1.

The first search sourced empirical research that provided examples of applying co-creation. The second search aimed to investigate gray literature to identify nonjournal publication types that provide guidance or information about co-creation methods, or include the use of co-creation methods.

### *Search 1: Empirical Research*

Health CASCADE, a European Commission-funded network that aims to develop a coherent methodology for co-creation research in public health, created a peer-reviewed curated database of co-creation literature containing 13,501 papers [29]. Since this study investigated methods used in co-creation, it was logical to source the methods from that database. However, since it contains a vast number of papers with different study types from various disciplines, a 2-step screening process was applied. First, a recursive search strategy was used to group the literature by method, reducing the number of papers for title and abstract screening. Second, title and abstract screening was conducted to categorize papers by study type.

The recursive search involved analyzing titles, abstracts, and keywords to identify relevant studies based on specific keywords, using Rayyan (Qatar Computing Research Institute), a systematic review manager [30]. This process included (1) screening titles and abstracts, (2) grouping by the method name, (3) screening grouped literature, and (4) iterating these steps until a stop rule was met. Details of this process are outlined in Multimedia Appendix 1.

Papers identified in the recursive search were extracted in Microsoft Excel and then reuploaded into Rayyan for classification by study type. We included empirical studies, protocols, exploratory studies, and case studies, while excluding evaluations and reflections on co-creation, to ensure the relevance of the extracted methods. This approach ensured that our inventory accurately reflects methods used in the co-creation process. For additional details, and the full set of inclusion and exclusion criteria, please refer to Multimedia Appendix 1. The classified literature was exported in a Microsoft Excel format and then taken to the analysis step.

Two approaches were used to analyze the final set of method names: one to deduce the frequency of the method names and one to validate the results. The first approach used pattern analysis, and the second approach used a more complex algorithm, commonly used by search engines. The pattern analysis provided initial frequency estimates, while the advanced algorithm validated these results by assessing the relevance of each method name in the context of the literature. This dual approach aimed to clarify which co-creation methods were most frequently used and which were less common. Further details can be found in Multimedia Appendix 1.

### *Search 2: Gray Literature Search*

An initial step in a systematic methods overview is identifying and selecting relevant literature, which can be poorly indexed in standard databases [25]. To address this, we conducted a search in Google's Advance Search tool for co-creation guidelines, toolkits, books, and nonjournal publications, as detailed in Multimedia Appendix 1.

Search strategies were aligned with those used to create the Health CASCADE Co-Creation Database covering literature dating back to January 1, 1970 [1]. We targeted specific domains and used co-creation and methods terms. The gray literature search strategy is outlined in Table 1 and provided in Multimedia Appendix 1.

**Table 1.** Gray literature search strategy.

| Parameter | Search 1 | Search 2 | Search 3 | Search 4 |
|---|---|---|---|---|
| Exact word or phrase | "co-creation methods" | "co-creation methods" | "co-creation methods" | co-creation AND methods AND guideline |
| Language | English | English | English | English |
| Region | Any region | Any region | Any region | Any region |
| Last updated | Anytime | Anytime | Anytime | Anytime |
| Date | January 1, 1970, to June 16, 2022 | January 1, 1970, to June 16, 2022 | any time | June 16, 2012, to June 16, 2022 |
| Domains | .org | .edu | .gov | .org |
| Terms appearing | Anywhere in the pages | Anywhere in the pages | Anywhere in the pages | Anywhere in the pages |
| File type | Any format | Any format | Any format | Any format |

We applied inclusion and exclusion criteria in 2 rounds to determine relevance. Initially, we included papers, reports, guidelines, books, and web-based tools relevant to co-creation, as defined by Agnello et al [1], provided they were absent from the Health CASCADE Co-Creation Database version 1.5 [1] and were written in English. Materials such as conference proceedings, abstracts, dissertations, non-English content, or those already included in the database were excluded. In the second round, selected full-text documents were rescreened to ensure they introduced and described at least 1 co-creation method, including its application or guidance on its use. Nonrelevant texts were excluded. The snowballing approach was used to trace citations and references of included literature to identify additional relevant resources, continuing until saturation was reached.

Literature that met all criteria was processed using a predefined Microsoft Excel extraction table. Data were validated by cross-checking the extracted method names with those found in both academic literature and gray literature, and the results were discussed among the coauthors (DMA, AS, and SC). Once all the extraction tables were completed, a list of methods and frequency of appearances across the included literature was generated, to ascertain the relevant prevalence of each method.

### Analysis

To analyze co-creation methods, we calculated method co-occurrence to understand how methods are combined in co-creation projects. The results were visualized using a Sankey diagram to illustrate the relationships between methods. We also conducted a bibliometric analysis of the included literature to investigate methodological trends and fields represented in the literature. Additionally, we examined the overlap between methods reported in academic literature and gray literature to determine if practitioners engage in co-creation similarly to those documented in academic research.

To understand how methods were combined, we analyzed co-occurrence in the academic literature by examining which methods appeared together in titles or abstracts. We used Python to map method names and count their co-occurrences, ensuring accuracy by marking and skipping already identified combinations. RAWGraphs (version 2.0; DensityDesign Research Lab) [31] was used to create Sankey diagrams that visualized these co-occurrences, removing method names with no co-occurrence. For ease of visualization, methods were categorized into three groups: (1) qualitative; (2) participatory; and (3) quantitative, mixed, and ethnographic. Separate Sankey diagrams for each group were produced to simplify and clarify the visualization of method combinations and their prevalence in co-creation research.

To gain insights into the methodological paradigms and fields within the academic literature, we performed a bibliometric analysis using VOSviewer (The Centre for Science and Technology Studies) [32]. This tool helped us construct and visualize bibliometric networks to assess the prevalence of different methodologies and research areas. By analyzing the source landscape and methodologies, we aimed to identify dominant fields and approaches represented in the included academic literature. Detailed steps for this analysis using VOSviewer are provided in Multimedia Appendix 2.

To determine if co-creation practices differ between academic and gray literature, we compared methods from both sources using a Venn diagram. We identified overlaps and unique methods by comparing the 2 lists of methods in Microsoft Excel using Conditional Formatting. Methods were categorized into three groups: (1) exclusively in academic literature, (2) exclusively in gray literature, and (3) found in both. We calculated the number and percentage of methods in each group and visualized these findings in a Venn diagram. This comparative analysis highlighted similarities and differences in co-creation approaches between academic and practitioner-focused literature.

## Results

### Systematic Methods Overview

#### *Search 1: Empirical Research*

For academic literature, the full set of literature from the Health CASCADE Co-Creation Database version 1.5—a total of 13,501 papers—were identified as relevant for screening. There were 2 subsequent screening processes for this literature: (1) grouping the papers by whether a method was present or not (n=13,501) and (2) screening the papers by study type (n=6472). During this 2-step screening process, a total of 10,905 papers were excluded, and 2590 papers were included.

The extraction of methods from academic literature involved 2 steps: first, method names were identified through a recursive search in Rayyan, and then, these names were used to determine their frequency in the final set of included literature. The identified methods were then subjected to co-occurrence analysis.

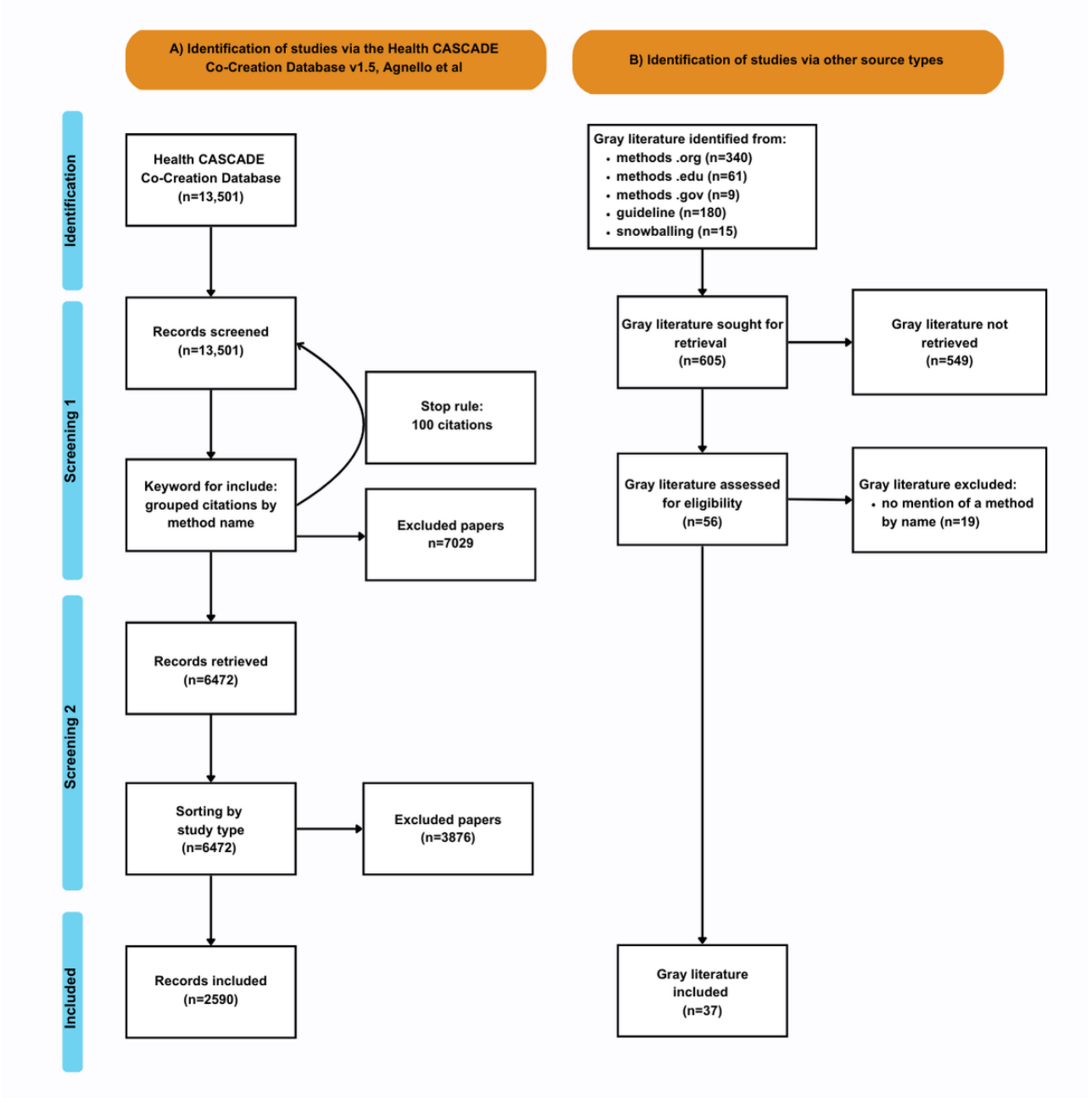
### *Search 2: Gray Literature Search*

A parallel search was conducted in gray literature, and a total of 605 materials were identified. As shown in Figure 1, there were two steps to screening the gray literature: (1) screening for relevance based on 1 set of selection criteria (n=56) and (2) screening the extracted full text based on a different set of selection criteria (n=37). These 2 screening steps resulted in the exclusion of 568 materials and the inclusion of 37 materials.

Method names from gray literature were manually extracted using a predetermined extraction form. The frequency of each method was calculated by counting its appearances across the included materials. For example, a method found in 5 different sources was assigned a frequency of 5.

A total of 2590 papers were included, and the gray literature search resulted in the inclusion of 37 materials. The adapted PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of this parallel process is shown in Figure 1, and a filled-in PRISMA checklist is provided in Multimedia Appendix 3.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the screening processes for the systematic methods overview with two sources as follows: (A) academic literature and (B) gray literature.



## Analysis

We assessed the relative presence of methods in the different types of literature by examining their frequency of occurrence. In academic sources, the frequencies ranged from 349 to 1, reflecting the varying specificity of method names. For example, “interview” appeared in 124 papers, while the more specific “Lego serious play” method was found in only 1 paper. Examples of the most and least frequent methods in academic literature are listed in Table 2, with the full set of methods and their frequencies provided in Multimedia Appendix 4.

In terms of gray literature, frequencies were manually tallied based on the extraction tables, ranging from 12 to 1. Examples of the most and least frequent methods sourced from gray literature are listed in Table 3, with the full set of methods and their relative frequencies provided in Multimedia Appendix 5.

Of the 252 methods extracted from academic literature, 91.3% (n=230) co-occurred within the same title or abstract. The resulting Sankey diagram visualizing this co-occurrence is too large to include in this manuscript and is therefore presented as an open-access figure on Zenodo [33], allowing readers to delve into the detailed co-occurrences and gain a deeper understanding of methods used in co-creation research. The full dataset of co-occurrence is provided in Multimedia Appendix 6. For clarity, the methods were categorized into three groups: (1) qualitative methods; (2) participatory methods; and (3) quantitative, mixed, and ethnographic methods, which are each visualized in separate Sankey diagrams.

**Table 2.** Examples of the most and least frequent methods that were sourced from academic literature. The percentage represents the percentage of total method hits.

| Method name | Values (n=3520), n (%) |
|---|---|
| **Most frequent methods** | |
| Survey | 349 (9.91) |
| Focus group | 337 (9.57) |
| Photo voice | 189 (5.37) |
| Group discussion | 150 (4.26) |
| Questionnaire | 142 (4.03) |
| Semistructured interview | 139 (3.95) |
| Interview | 124 (3.52) |
| **Least frequent methods** | |
| Consensus workshop | 1 (0) |
| Social mapping | 1 (0) |
| Participatory theme elicitation | 1 (0) |
| Lego serious play | 1 (0) |
| User persona | 1 (0) |
| Emotional touchpoints | 1 (0) |
| Structured brainstorm | 1 (0) |

**Table 3.** Examples of the most and least frequent methods that were sourced from gray literature. The percentage represents the percentage of total extracted method hits.

| Method name | Values (n=1151), n (%) |
|---|---|
| **Most frequent methods** | |
| World café | 12 (1) |
| Focus group | 10 (1) |
| Role playing | 9 (1) |
| Persona | 8 (1) |
| Brainstorming | 7 (1) |
| Card sorting | 7 (1) |
| Storyboarding | 7 (1) |
| **Least frequent methods** | |
| What if brainstorming | 1 (0) |
| Trigger storming | 1 (0) |
| The blue sky vision exercise | 1 (0) |
| System mapping | 1 (0) |
| Sorting important to/for | 1 (0) |
| Sky the limit brainstorm | 1 (0) |
| Service safari | 1 (0) |

The co-occurrence analysis of the qualitative methods, shown in Figure 2, features 22 source methods (qualitative) on the left and 149 target methods (multiple types) on the right side of the Sankey diagram. This analysis revealed that focus group, one of the most frequently used methods, often co-occurs with other qualitative methods like group discussion, interviews, and in-depth interviews, as well as qualitative analysis such as content analysis and thematic analysis. Additionally, focus group commonly co-occurs with participatory methods like prototyping, photo voice, storytelling, and various ethnographic methods, including participant observation, field notes, and narrative.

The co-occurrence analysis of the participatory methods, visualized in Figure 3, includes 42 source methods (participatory) on the left and 132 target methods (multiple types) on the right side of the Sankey diagram. This analysis highlighted that some of the most frequently used methods, such as photo voice, and prototyping, often co-occur with other participatory methods like experience prototyping, concept mapping, and participatory mapping. Photo voice is also commonly linked to narrative and group discussion, and a mix of ethnographic and qualitative methods, including thematic analysis, field observation, and various types of interviews. Additionally, deliberative workshops frequently co-occur with other deliberative or participatory methods such as user committee, participatory budgeting, fuzzy cognitive mapping, and collective reflection.

The co-occurrence analysis of quantitative, mixed, and ethnographic methods is presented in Figure 4, with 24 source methods (quantitative, mixed, and ethnographic) on the left and 156 target methods (multiple types) on the right side of the Sankey diagram. This analysis showed that the most frequently used method, survey, often co-occurs with other high-frequency methods like questionnaire and narrative. Survey is also strongly linked to qualitative methods such as focus groups, group discussions, various types of interviews, and qualitative analysis methods. The ethnographic method, field notes, is associated with a combination of ethnographic and participatory methods, including participatory observation, participatory reflection, and group model building.

The VOSviewer co-occurrence analysis of keywords related to methodologies in the title and abstracts of the academic literature is visualized as a network map in Figure 5. This map includes 34 keywords, interconnected a total of 385 times, and organized by VOSviewer into 4 clusters [32]. Cluster 1 (in yellow) includes terms related to participatory research, community-based participatory research, and community engagement or participation. Cluster 2 (in teal) focuses on various forms of co-creation, value co-creation, and design methodology approaches. Cluster 3 (in pink) includes coproduction, co-design, public and patient involvement, and user involvement. Cluster 4 (in blue) groups together different forms of participatory action research and action research. The main terms across these 4 clusters are co-creation, coproduction, co-design, participatory action research, and participatory research. Further details about these clusters are provided in Table 4.

**Figure 2.** Sankey diagram illustrating the co-occurrence of qualitative methods with other research methods in academic literature. Source methods are displayed on the left side (n=22) and target methods are on the right (n=149). The thickness of the line indicates the frequency of co-occurrences across the literature, with thicker lines representing more frequent co-occurrences between methods.

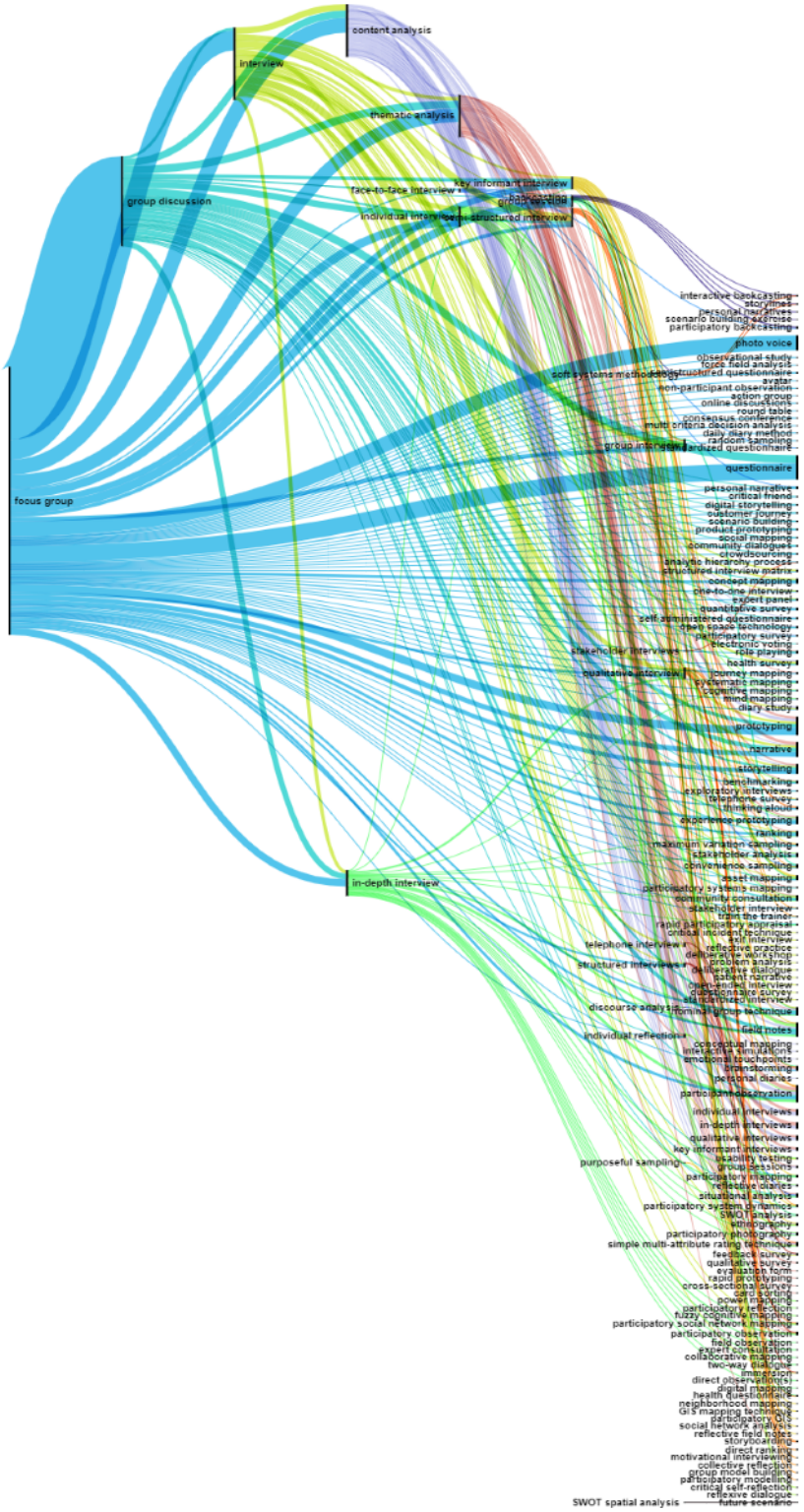


**Figure 3.** Sankey diagram illustrating the co-occurrence of participatory methods with other research methods in academic literature. Source methods are displayed on the left side (n=42) and target methods are on the right (n=132). The thickness of the line indicates the frequency of co-occurrences across the literature, with thicker lines representing more frequent co-occurrences between methods.

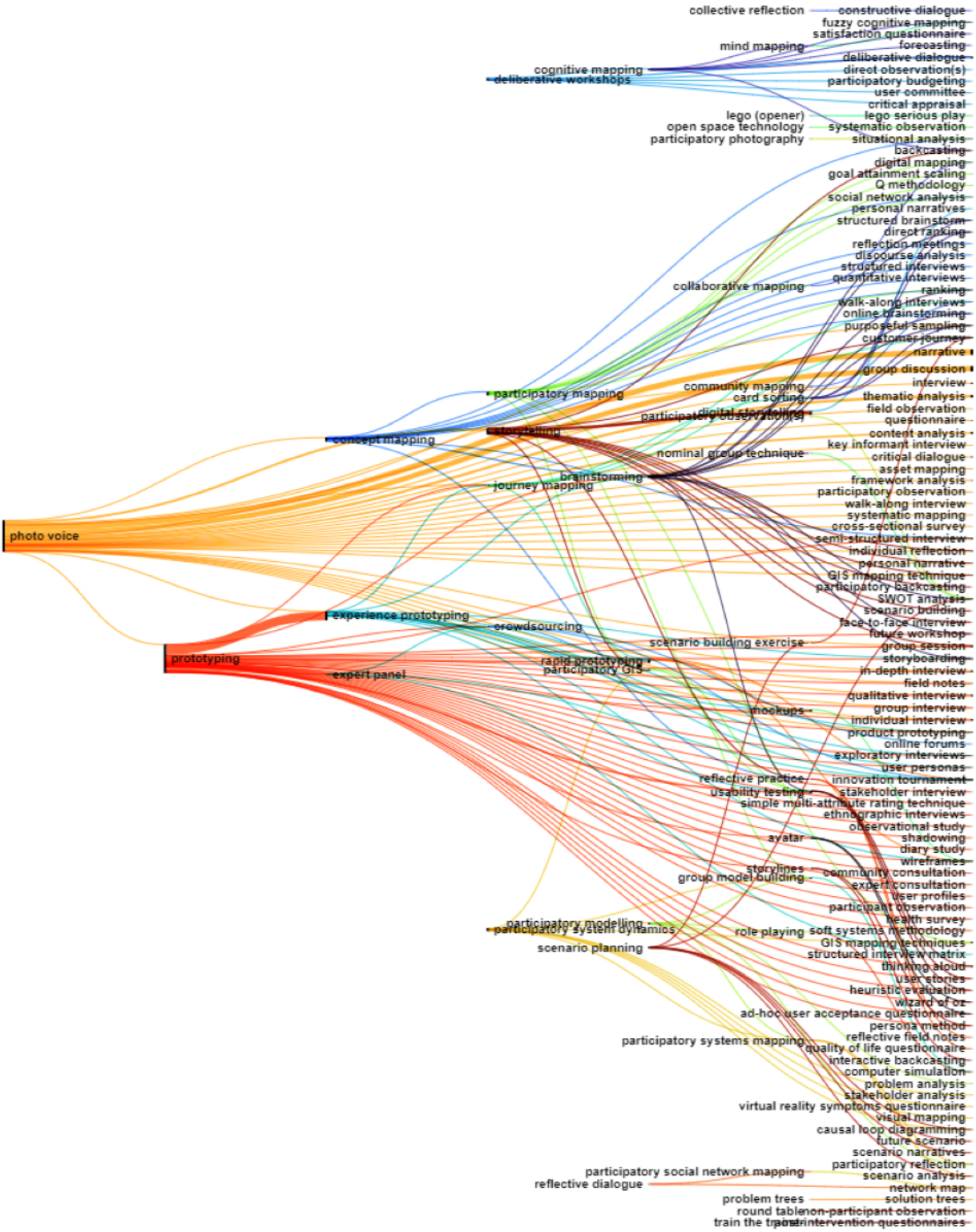


**Figure 4.** Sankey diagram illustrating the co-occurrence of quantitative, mixed, and ethnographic methods with other research methods in academic literature. Source methods are displayed on the left side (n=24) and target methods are on the right (n=156). The thickness of the line indicates the frequency of co-occurrences across the literature, with thicker lines representing more frequent co-occurrences between methods.

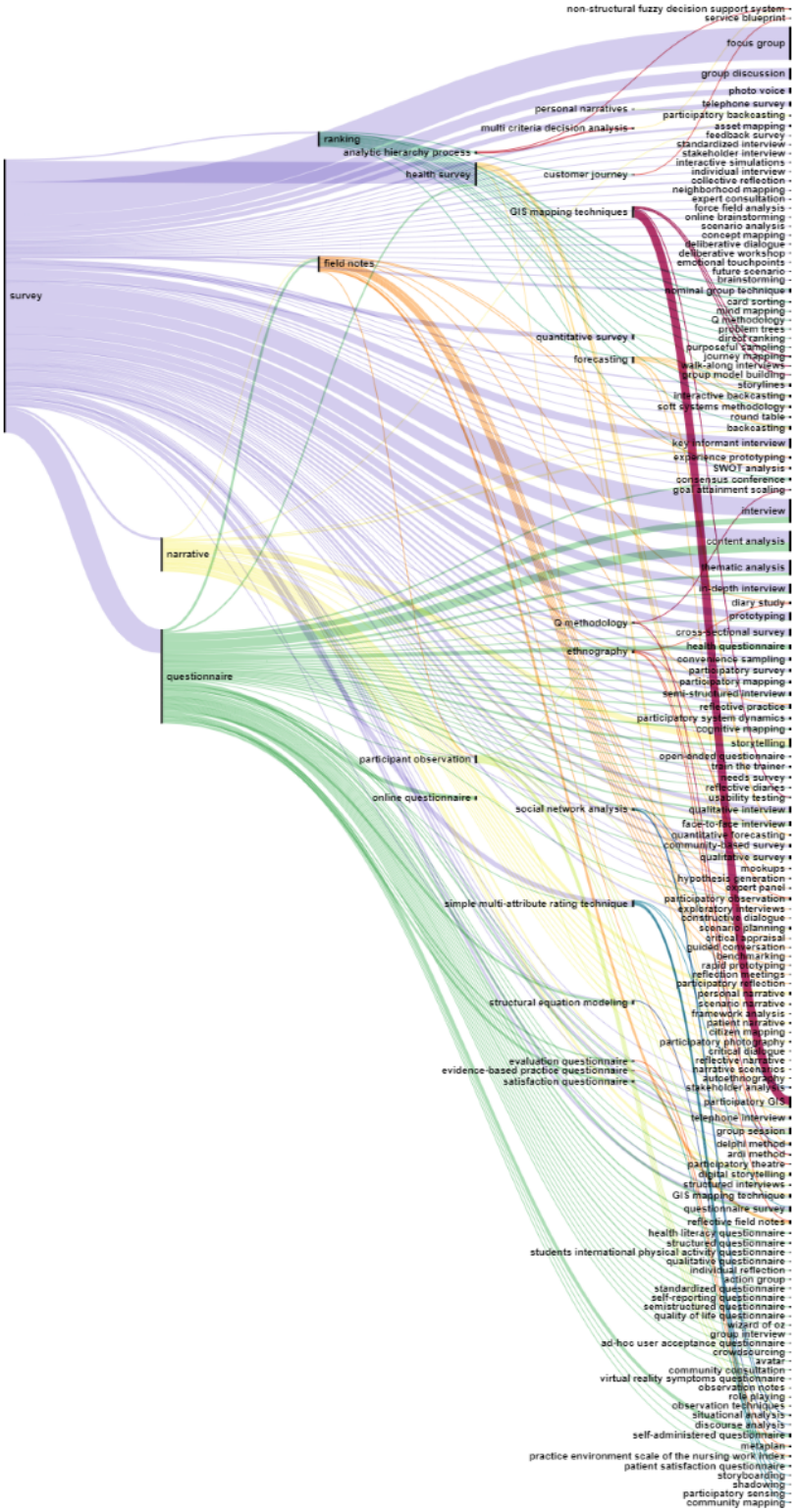


**Figure 5.** VOSviewer generated an image representing the co-occurrence of keywords representing methodologies across the title and abstracts of the academic literature. The size of the keyword bubble represents its importance in the number of co-occurrences. Each line represents a co-occurrence of the terms. cbpr: community-based participatory research; par: participatory action research; ppi: public and patient involvement.

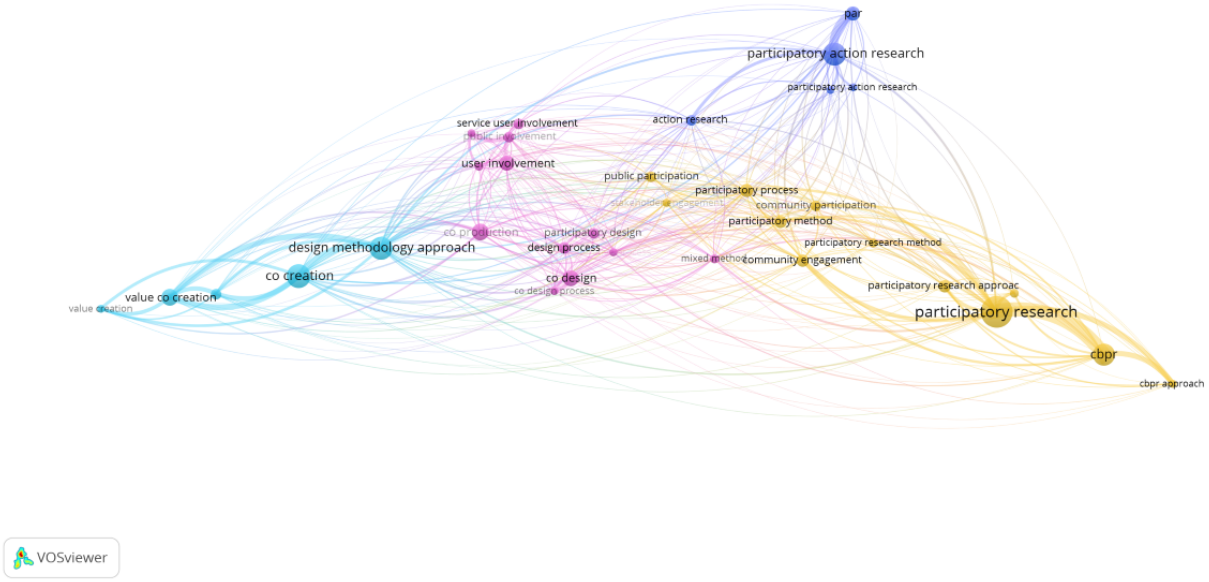



**Table 4.** Co-occurrence of methodologies in academic literature.

| Methodology (Cluster) | Co-occurrence links (n=385), n (%) | Related methodologies |
|---|---|---|
| Participatory research (yellow) | 31 (8) | Participatory process, participatory method, participatory research approach, participatory research method, participatory research project, community-based participatory research, community-based participatory research approach, public participation, stakeholder engagement, community engagement, and community participation. |
| Co-creation (teal) | 29 (8) | Co-creation process, design methodology approach, value co-creation, and value creation. |
| Coproduction (pink) | 28 (7) | Co-design, co-design process, design process, iterative process, mixed method, participatory design, patient involvement, public and patient involvement, public involvement, and user involvement. |
| Participatory action research (blue) | 25 (7) | Action research, participatory action research, participatory action research project, and participatory action research approach. |

The VOSviewer co-occurrence analysis of the keywords related to fields in the title and abstracts of papers in the database is visualized as a network map in Figure 6. This map includes 150 keywords, interconnected 181 times and organized by VOSviewer into 48 clusters [32]. Focusing on the 8 most prominent clusters, which contain at least 5 keywords each, cluster 1 (in red) includes health service research, community health services, medical education, occupational therapy, primary health care, and social work. Cluster 2 (in blue) covers health promotion, health policy, and school health services. Cluster 3 (in green) includes dementia, emergency nursing, home care services, home nursing, long-term care, and terminal care. Cluster 4 (in purple) features palliative care, emergency medical services, and hospital emergency services. Cluster 5 (in orange) focuses on women's health, maternal welfare, rural health, and supported employment. Cluster 6 (in yellow) includes humanities, pediatric hospitals, organizational innovation, quality health care, and vaccinations. Cluster 7 (in brown) is centered on health and nursing education. Cluster 8 (in teal) covers family health, child development disorders, learning disorders, and self-evaluation programs. Full details on these clusters are provided in Table 5.

This study extracted a total of 956 methods used in co-creation, which were then categorized into 3 distinct categories based on their sources. Of these, 16% (155/956) were found exclusively in academic literature, 74% (704/956) were sourced solely from gray literature, and 10% (97/956) were identified in both academic literature and gray literature. The comparison of these categories is visualized in Figure 7, and a complete list of methods per source type is provided in Multimedia Appendix 7.

**Figure 6.** VOSviewer generated an image representing the co-occurrence of keywords representing fields across the titles and abstracts of the academic literature. The size of the keyword bubble represents its importance in the number of co-occurrences. Each line represents a co-occurrence of the terms.

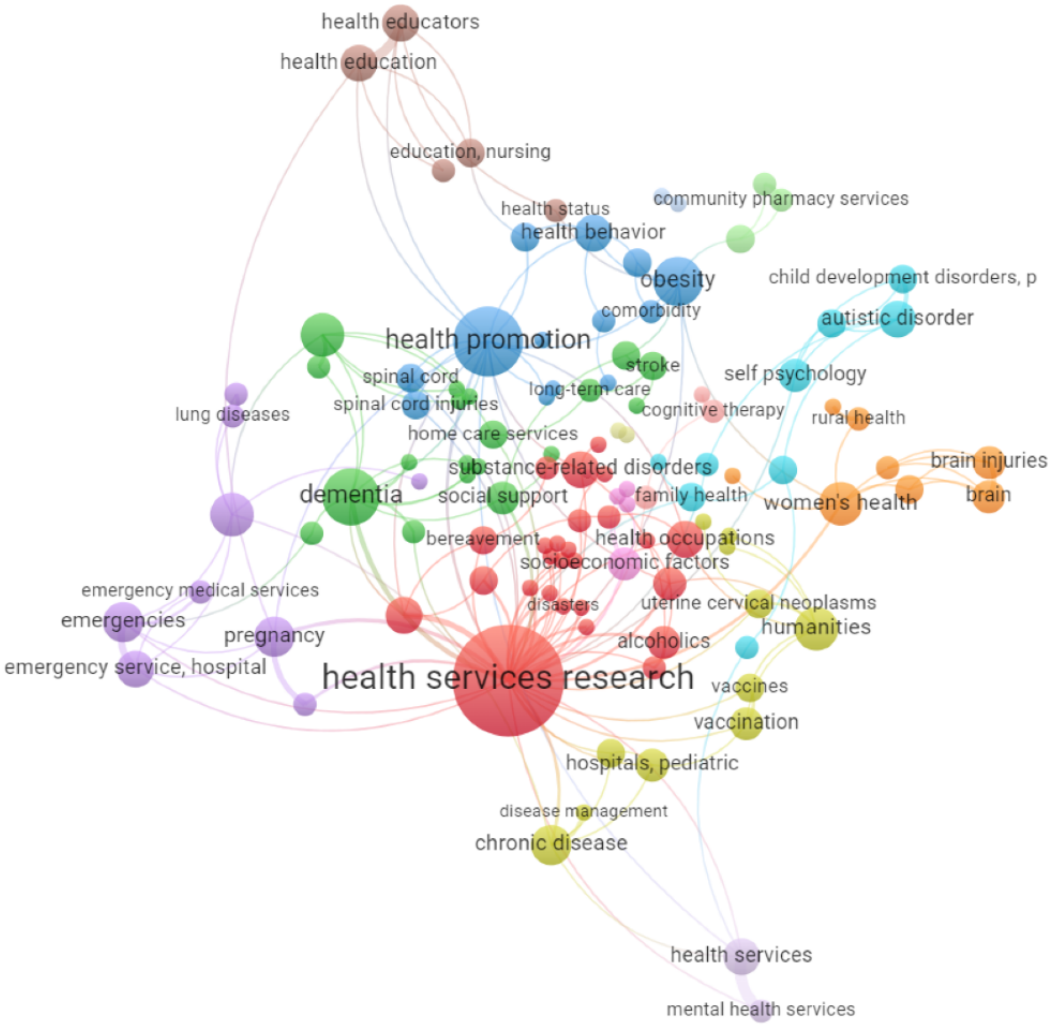
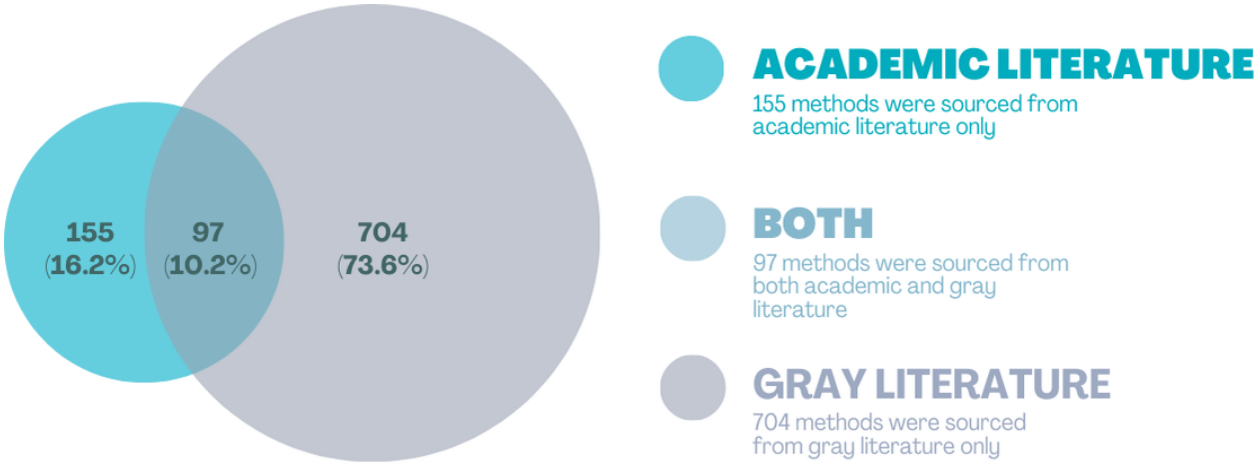



**Table 5.** Co-occurrence of fields and diseases in academic literature.

| Field (cluster) | Co-occurrence links (n=181), n (%) | Related topics |
|---|---|---|
| Health services research (red) | 46 (25) | Community health services, medical education, health service accessibility, occupational health, occupational therapy, primary health care, social work, and substance-related disorders. |
| Health promotion (blue) | 18 (10) | Health policy, mental health, professional practice, school health services, rheumatic diseases, and spinal cord injuries. |
| Dementia (green) | 12 (7) | Emergency nursing, home care services, home nursing, long-term care, community health nurses, social support, and terminal care. |
| Palliative care (purple) | 7 (4) | Emergencies, emergency medical services, hospital emergency services, pregnancy, and lung diseases. |
| Women’s health (orange) | 7 (4) | Brain injuries, maternal welfare, rural health, and supported employment. |
| Humanities (yellow) | 7 (4) | Disease management, pediatric hospitals, chronic disease, organizational innovation, quality health care, and vaccinations. |
| Health education (brown) | 5 (3) | Nursing education, health educators, and health status. |
| Family health (teal) | 3 (2) | Anxiety, autistic disorder, child development disorder, disabled persons, learning disorders, and self-evaluation programs. |

**Figure 7.** Venn diagram of methods sourced from academic, gray literature, or both. Percentages represent the portion of the total methods (n=956).




## Discussion

### Principal Findings

This study is the first systematic methods overview of co-creation, identifying and delving into both academic and gray literature to extract 956 methods from 2627 sources, analyzing these methods and their interrelationships within and across source types. This study can facilitate the identification of methods suitable for co-creation endeavors, while also highlighting discrepancies between co-creation in research and nonacademic contexts. By offering a comprehensive and reliable snapshot of co-creation and its associated methods, this research enhances transparency regarding how co-creation is executed in both research and practice in the health research fields. The study serves as a foundational resource, providing an evidence-based and systematic inventory of methods, as well as shedding light on which methods are commonly used in conjunction with one another across the health research fields. This innovative approach to conducting a systematic methods overview enables the comprehensive analysis of a vast amount of literature and methods, offering valuable insights into the landscape of co-creation practices and methods globally and across disciplines.

### Principal Results

A notable strength of this study lies in its utilization of methods sourced from a published and high-quality database of co-creation literature spanning from 1970 to 2022. This literature represents the largest known dataset of methods used in co-creation. This study serves as a foundational work that can be used to elucidate the interrelationships between different co-creation methods and co-creation processes. This study adopts an unbiased approach by providing a comprehensive overview of current co-creation practices, without prescribing specific methods or recommendations. By providing an overview of current co-creation practices, this study establishes a platform for further research and analysis into the various ways in which people design and execute co-creation. Furthermore, this study reveals a stark difference between academic and practitioner approaches to co-creation, highlighting the importance of understanding and bridging these disparities.

### Multimethod Approach

In the realm of research methodologies, each type has its unique characteristics and applications. Recently, Messiha et al [11] discussed how naïve realism (a singular objective reality) often favors quantitative methods, while relativism (multiple realities exist) tends to favor qualitative approaches. For instance, approaches such as qualitative, quantitative, and ethnographic studies offer different lenses through which to explore phenomena. Qualitative methods delve into the subjective experiences and perspectives of individuals, capturing rich narratives and deep insights; valuing methodological pluralism and diversity rather than universal truths [34]. Quantitative methods, on the other hand, focus on numerical data and statistical analysis to uncover patterns and relationships to uncover universal truths about reality. Finally, participatory methods emphasize collaboration between researchers and stakeholders; unlike traditional research methods, which prioritize objectivity, participatory methods prioritize engagement and collective decision-making [27]. Furthermore, Messiha et al [35] found that applying methodological principles of critical realism (recognizes independent reality and acknowledges multiple perspectives), namely using multiple method types to understand complex phenomena, enriched the evidence base for co-creation in public health research.

This study reveals that co-creation represents a fusion of various types of methods, and well-known methods for analysis, such as thematic analysis, are combined with qualitative methods like interviews. Co-creation combines methods for qualitative inquiry and analysis, quantitative analysis, ethnographic observation, and participatory collaboration. This amalgamation of methods offers great potential for generating innovative solutions, fostering stakeholder engagement, and addressing complex challenges. However, the convergence of multiple methods also introduces challenges, including potential divergences in processes and outcomes between co-creation projects. Overall, the mixed nature of co-creation presents both opportunities and challenges.

### The Gap Between Academics and Practitioners

While analyzing the data within this study, a clear gap exists between the practices reported in academic literature and those used by practitioners in nonacademic settings. This divergence

may stem from various factors, including differences in reporting styles, publication platforms, and the emphasis placed on outcomes versus process documentation. Academic publications often prioritize the reporting of research outcomes and their impact, allocating limited space for detailing the intricacies of the co-creation process. Conversely, practitioners may opt to disseminate their co-creation endeavors through alternative channels, such as reports, case studies, or web-based platforms, where they have more flexibility to document their methods. However, this discrepancy in reporting practices poses a significant challenge for the scientific community, which is well documented by various researchers who are calling for researchers to transparently document the entire process of co-creation, including the methods [6,14,15]. Without detailed documentation of each method used in the co-creation process, replication and validation of findings become challenging. Furthermore, the lack of transparency in documenting co-creation methods hinders knowledge sharing and collaboration among researchers and practitioners. This gap underscores the importance of adopting standardized reporting guidelines and best practices for documenting co-creation, regardless of the publication platform.

One notable observation is the prevalence of co-creation methods outside of academic literature. While academic research may provide valuable insights into academic rigor, and application of theory, it appears that the majority of the co-creation methods are published on nonacademic platforms. This raises questions about the transparency and documentation of co-creation processes. Efforts to bridge the divide between academic and practitioner perspectives on co-creation are essential for advancing knowledge, promoting collaboration, and maximizing the impact of co-creation initiatives. By fostering greater transparency, documentation, and knowledge sharing, researchers and practitioners can enhance the rigor and reproducibility of co-creation.

## Comparison With Prior Work

Grindell et al [16] underscore the distinctiveness of collaborative approaches like co-creation, co-design, and coproduction, contrasting them with traditional applied health research methodologies; emphasizing that these coapproaches are designed to foster meaningful engagement with a broader range of cocreators, including those not typically involved in research. However, the network analysis of the methodologies in the final dataset reveals a different story. While coapproaches and participatory research methodologies are prevalent, the actual methods used in these projects fall short of truly embracing collaborative engagement. This highlights a disconnect between the intended methodological goals and the practical application of collaborative methods.

Our findings align with recent work by Slattery et al [14] and Smith et al [15], who reported similar trends in the types of methods being used. Slattery et al [14] observed that the most frequently used methods were focus groups, interviews, and surveys, with less frequent use of methods like citizen juries and voting. Smith et al [15] also found that commonly reported methods included interviews, focus group discussions, surveys, discussion meetings, and workshops. This consistency across studies is encouraging, as it underscores a broader trend in the field and validates our findings about the dominant reliance on certain qualitative methods.

Louise and Annette [12] highlight the significance of using participatory methods, emphasizing their potential applicability to intervention development and their capacity to transcend traditional boundaries. When executed effectively, participatory methods offer numerous benefits, including genuine stakeholder involvement, a holistic understanding of multiple perspectives, an iterative and investigative approach, and a commitment to enacting meaningful change for those directly affected by the outcomes [12]. Maenhout et al [6] also found that co-creation is feasible with adolescents with intellectual disabilities if the correct methods are selected, specifically creative methods. Furthermore, studies favoring creative methods found that design and participatory approaches effectively engaged cocreators with their emotions, and abilities, and retained their involvement throughout the process [16,36]. Prototyping methods were found to be valuable for translating knowledge into tangible objects, while visual design methods facilitated the rapid communication of ideas in an accessible manner. The utilization of creative methods was observed to promote a shared understanding of the problem and identify critical needs, thus addressing power differentials and fostering a sense of ownership among stakeholders [16].

Grindell et al [16] illustrated how researchers rely on qualitative research methods (such as focus groups, observations, and interviews) when other, more creative methods, can achieve the same aim (such as role-playing, personas, and user journeys). These findings underscore the importance of using participatory and creative approaches in co-creation processes to build trust, confidence, and collaborative solutions. Considering the documented benefits of using these types of methods, this study reveals that the way people are currently cocreating is leaning heavily on qualitative methods. While qualitative methods provide rich insights into individual subjective experiences and perspectives, incorporating diverse approaches such as creative or participatory methods could further enhance stakeholder engagement, understanding, and problem-solving in co-creation processes.

There is a growing concern regarding participation fatigue in co-creation, alongside the risk of heightened inequity when only specific demographic groups engage, potentially leading to disengagement [37,38]. Amundrud et al [38] stress that co-creation should empower individuals to participate actively as citizens with the chance to genuinely influence outcomes. Various methods, including focus groups, one-on-one discussions, and creative workshops, are utilized for involvement and engagement. However, patients increasingly advocate for greater autonomy in determining the extent and nature of their involvement in research processes. In contexts like coproduction and participatory action research, traditional distinctions between researchers and participants are challenged, resulting in blurred boundaries and a reassessment of power dynamics [12]. This study reveals that the multi-method approach to co-creation primarily integrates various qualitative methods, while more creative or participatory methods are often found in nonacademic gray literature. Consequently, academics appear

to rely solely on qualitative methods and rarely draw from nonacademic sources for guidance and inspiration. This heavy reliance on qualitative methods may increase the risk of disengagement, reassessment of power dynamics, or participation fatigue.

### Consequences and Future Research

This systematic inventory of methods used in co-creation provides an interesting dataset that should be mined, for instance, to make further distinctions between the methods and their intended purpose and impact. We intend to use it to develop a taxonomy facilitating more consistent and better reporting, which other reviews highlighted as an important gap [14,15]. In turn, this could facilitate a more robust evaluation of the impact of co-creation, which is fundamental for the progression of this methodology [15,16]. Co-creation researchers could make use of such a taxonomy or apply our novel artificial intelligence and data science approach, to extract additional information about the methods used in co-creation, mirroring the approach of the Human Behavior Project for behavior change science [39]. This approach could help identify specific clusters of methods or sequences of methods that are particularly effective in certain contexts, for certain purposes, and in certain phases of the co-creation process. The data generated in this study also raises interesting questions about how and when collective intelligence and creative methods are best placed or used, and how scientific practices might have to change to provide room for effective collective creativity based on evidence.

### Limitations

Although we achieved our objectives by conducting a systematic overview of methods, our approach has some limitations including single screening of the gray literature and the lack of full-text screening of the academic literature. However, we recognize that this work was conducted in a peer-reviewed, precurated database of co-creation literature, and the complexity of the screening of gray literature was not high enough to necessitate double screening and extraction.

### Conclusions

This study has yielded the first systematic overview of methods utilized in co-creation, intending to provide researchers and practitioners with a comprehensive understanding of methods used in co-creation to date. By increasing awareness of methods used in co-creation, we seek to unlock the full potential of co-creation and contribute to its advancement as a transformative methodology for research and practice. The analysis of methods sourced from both academic and gray literature revealed a rich array of methods used in co-creation, spanning participatory, qualitative, quantitative, mixed, and ethnographic approaches. This diversity underscores the versatility of methods, which can adapt to varying study objectives, target groups, contexts, and other influencing factors. However, it also highlights the need for more detailed guidance on method selection, method grouping, and application to ensure co-creation remains effective and meaningful.

This systematic exploration of co-creation methods offers valuable insights for individuals currently involved in co-creation, as well as those aspiring to participate in it. By illuminating the diverse landscape of co-creation methods, this study aims to enable researchers and practitioners to make informed decisions and enhance methodological rigor and innovation in co-creation. Through continued research and collaboration, we can further advance co-creation as a dynamic and impactful approach for addressing some of the most pressing complex or wicked public health challenges.

## Acknowledgments

The authors would like to thank Julie Smith, Librarian at Glasgow Caledonian University, for her input on the gray literature search strategy. This study was funded by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement 956501. The views expressed in this manuscript are the author's views and do not necessarily reflect those of the funders.

## Data Availability

The inventory of methods sourced from both academic and gray literature is available on Zenodo under the name "Co-Creation Methods Inventory: Sourced from Academic and Grey Literature" [40]. The Sankey diagram is available on Zenodo under the name "Co-occurrence of Methods in Co-Creation: A Sankey Diagram" [33].

## Authors' Contributions

DMA, AS, and SC conceived the study, and DMA led the study. DMA wrote the manuscript. AS, DMA, GB, and SC edited the manuscript. DMA developed the search strategy. DMA conducted the search in academic and gray literature. DMA executed the screening steps, and extraction, conducted the bibliometric analysis, and generated all the figures. GB conducted the method frequency and co-occurrence analysis. The manuscript was then reviewed and approved by all authors.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Extended methods description.
[DOCX File , 230 KB - ijmr_v13i1e59772_app1.docx ]

Multimedia Appendix 2
VOSviewer analysis steps.
[DOCX File , 25 KB - ijmr_v13i1e59772_app2.docx ]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 33 KB - ijmr_v13i1e59772_app3.docx ]

Multimedia Appendix 4
Academic literature methods.
[XLSX File (Microsoft Excel File), 17 KB - ijmr_v13i1e59772_app4.xlsx ]

Multimedia Appendix 5
Gray literature extraction.
[XLSX File (Microsoft Excel File), 78 KB - ijmr_v13i1e59772_app5.xlsx ]

Multimedia Appendix 6
Co-occurrence Sankey data.
[XLSX File (Microsoft Excel File), 29 KB - ijmr_v13i1e59772_app6.xlsx ]

Multimedia Appendix 7
Academic versus gray literature.
[XLSX File (Microsoft Excel File), 32 KB - ijmr_v13i1e59772_app7.xlsx ]

## References


1. Agnello DM, Loisel QEA, An Q, Balaskas G, Chrifou R, Dall P, et al. Establishing a health CASCADE-curated open-access database to consolidate knowledge about co-creation: novel artificial intelligence-assisted methodology based on systematic reviews. J Med Internet Res 2023;25:e45059 [FREE Full text] [doi: 10.2196/45059] [Medline: 37463024]
2. Darlington E, Masson J. What does co-creation mean? an attempt at definition informed by the perspectives of school health promoters in France. Health Educ J 2021;80(6):746-758. [doi: 10.1177/00178969211013570]
3. van Woezik AFG, Braakman-Jansen LMA, Kulyk O, Siemons L, van Gemert-Pijnen JEWC. Tackling wicked problems in infection prevention and control: a guideline for co-creation with stakeholders. Antimicrob Resist Infect Control 2016;5:20 [FREE Full text] [doi: 10.1186/s13756-016-0119-2] [Medline: 27213040]
4. Greenhalgh T, Jackson C, Shaw S, Janamian T. Achieving research impact through co-creation in community-based health services: literature review and case study. Milbank Q 2016;94(2):392-429 [FREE Full text] [doi: 10.1111/1468-0009.12197] [Medline: 27265562]
5. Lewis CC, Taba M, Allen TB, Caldwell PH, Skinner SR, Kang M, et al. Developing an educational resource aimed at improving adolescent digital health literacy: using co-design as research methodology. J Med Internet Res 2024;26:e49453 [FREE Full text] [doi: 10.2196/49453] [Medline: 39110967]
6. Maenhout L, Verloigne M, Cairns D, Cardon G, Crombez G, Melville C, et al. Co-creating an intervention to promote physical activity in adolescents with intellectual disabilities: lessons learned within the move it, move ID!-project. Res Involv Engagem 2023;9(1):10 [FREE Full text] [doi: 10.1186/s40900-023-00420-x] [Medline: 36935503]
7. Mallakin M, Dery C, Vaillancourt S, Gupta S, Sellen K. Web-based co-design in health care: considerations for renewed participation. Interact J Med Res 2023;12:e36765 [FREE Full text] [doi: 10.2196/36765] [Medline: 36595738]
8. Wong CC, Kumpulainen K, Kajamaa A. Collaborative creativity among education professionals in a co-design workshop: a multidimensional analysis. Think Ski Creat 2021;42:100971. [doi: 10.1016/j.tsc.2021.100971]
9. Wise S, Paton R, Gegenhuber T. Value co-creation through collective intelligence in the public sector: a review of US and European initiatives. VINE 2012:251-276. [doi: 10.1108/03055721211227273]
10. Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW. Evidence for a collective intelligence factor in the performance of human groups. Science 2010;330(6004):686-688. [doi: 10.1126/science.1193147] [Medline: 20929725]
11. Messiha K, Chinapaw M, Ket HCFF, An Q, Anand-Kumar V, Longworth GR, et al. Systematic review of contemporary theories used for co-creation, co-design and co-production in public health. J Public Health (Oxf) 2023;45(3):723-737 [FREE Full text] [doi: 10.1093/pubmed/fdad046] [Medline: 37147918]

12. Louise L, Annette B. Drawing straight lines along blurred boundaries: qualitative research, patient and public involvement in medical research, co-production and co-design. evid policy 2019;15(3):409-421. [doi: 10.1332/174426419X15552999451313]
13. An Q, Sandlund M, Agnello D, McCaffrey L, Chastin S, Helleday R, et al. A scoping review of co-creation practice in the development of non-pharmacological interventions for people with chronic obstructive pulmonary disease: a health CASCADE study. Respir Med 2023;211:107193 [FREE Full text] [doi: 10.1016/j.rmed.2023.107193] [Medline: 36889517]
14. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
15. Smith H, Budworth L, Grindey C, Hague I, Hamer N, Kislov R, et al. Co-production practice and future research priorities in United Kingdom-funded applied health research: a scoping review. Health Res Policy Syst 2022;20(1):36 [FREE Full text] [doi: 10.1186/s12961-022-00838-x] [Medline: 35366898]
16. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022;22(1):877 [FREE Full text] [doi: 10.1186/s12913-022-08079-y] [Medline: 35799251]
17. Quintão C, Andrade P, Almeida F. How to improve the validity and reliability of a case study approach? J Interdiscip Stud Educ 2020;9(2):273-284. [doi: 10.32674/jise.v9i2.2026]
18. Osuagwu LC. Research methods: issues and research direction. Bus Manag Res 2020;9(3):46. [doi: 10.5430/bmr.v9n3p46]
19. Dubé L, Paré G. Rigor in information systems positivist case research: current practices, trends, and recommendations. MIS Q 2003;27(4):597. [doi: 10.2307/30036550]
20. Leung L. Validity, reliability, and generalizability in qualitative research. J Family Med Prim Care 2015;4(3):324-327 [FREE Full text] [doi: 10.4103/2249-4863.161306] [Medline: 26288766]
21. Lee JJ, Jaatinen M, Salmi A, Mattelmäki T, Smeds R, Holopainen M. Design choices framework for co-creation projects. Int J Des 2018;12(2):15-31 [FREE Full text]
22. Durugbo C, Pawar K. A unified model of the co-creation process. Expert Syst Appl 2014;41(9):4373-4387. [doi: 10.1016/j.eswa.2014.01.007]
23. Stone-Romero EF. The relative validity and usefulness of various empirical research designs. In: Rogelberg SG, editor. Handbook of Research Methods in Industrial and Organizational Psychology. Oxford, UK: Blackwell Publishing; 2002:77-98.
24. Patten ML. Understanding Research Methods: An Overview of the Essentials. 10th ed. New York, NY: Routledge; 2017.
25. Gentles SJ, Charles C, Nicholas DB, Ploeg J, McKibbon KA. Reviewing the research methods literature: principles and strategies illustrated by a systematic overview of sampling in qualitative research. Syst Rev 2016;5(1):172 [FREE Full text] [doi: 10.1186/s13643-016-0343-0] [Medline: 27729071]
26. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot Feasibility Stud 2019;5:41 [FREE Full text] [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
27. Vaughn L, Jacquez F. Participatory research methods – choice points in the research process. J Particip Res Methods 2020;1(1). [doi: 10.35844/001c.13244]
28. Gentles SJ, Vilches SL. Calling for a shared understanding of sampling terminology in qualitative research: proposed clarifications derived from critical analysis of a methods overview by McCrae and Purssell. Int J Qual Methods 2017;16(1):160940691772567. [doi: 10.1177/1609406917725678]
29. Loisel Q, Agnello D, Chastin S. Co-creation database. zenodo 2022. [doi: 10.5281/zenodo.6773028]
30. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
31. RAWGraphs 2.0. URL: https://app.rawgraphs.io/ [accessed 2024-03-20]
32. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
33. Agnello D, Balaskas G, Steiner A, Chastin S. Co-occurrence of methods in co-creation: a sankey diagram. Zenodo 2024. [doi: 10.5281/zenodo.10978194]
34. Parker L. Qualitative perspectives: through a methodological lens. Qual Res Account Manag 2014;11(1):13-28. [doi: 10.1108/QRAM-02-2014-0013]
35. Messiha K, Altenburg TM, Schreier M, Longworth GR, Thomas N, Chastin S, et al. Enriching the evidence base of co-creation research in public health with methodological principles of critical realism. Critical Public Health 2024;34(1):1-19. [doi: 10.1080/09581596.2024.2371323]
36. Longworth GR, Erikowa-Orighoye O, Anieto EM, Agnello DM, Zapata-Restrepo JR, Masquillier C, et al. Conducting co-creation for public health in low and middle-income countries: a systematic review and key informant perspectives on implementation barriers and facilitators. Global Health 2024;20(1):9. [doi: 10.1186/s12992-024-01014-2] [Medline: 38233942]
37. Pappers J, Keserü I, Macharis C. Co-creation or public participation 2.0? an assessment of co-creation in transport and mobility research. In: Towards User-Centric Transport in Europe 2. Lecture Notes in Mobility. Cham, Switzerland: Springer International Publishing; 2020.

38. Amundrud A, Christensen I, Smørdal O, Ludvigsen S. Exploring the potential of a co-creation platform for children and youth. 2023 Presented at: 16th International Conference on Computer-Supported Collaborative Learning - CSCL 2023; June 10-11, 2023; Montréal, QC p. 392-393 URL: https://repository.isls.org//handle/1/9247 [doi: 10.22318/cscl2023.500423]
39. Welcome to the human behaviour change project including the APRICOT project. Human Behaviour Change Project. URL: https://www.humanbehaviourchange.org/ [accessed 2024-08-21]
40. Agnello D. Co-creation methods inventory: sourced from academic and grey literature. Published online July 21 2023. [doi: 10.5281/zenodo.8014373]

## Abbreviations

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

*Edited by Z Guo; submitted 23.04.24; peer-reviewed by C Kunze, K Ryom, A Davis, J Downey, E Avram, M Schreier; comments to author 14.07.24; revised version received 06.09.24; accepted 02.10.24; published 11.11.24.*

*Please cite as:*
*Agnello DM, Balaskas G, Steiner A, Chastin S*
*Methods Used in Co-Creation Within the Health CASCADE Co-Creation Database and Gray Literature: Systematic Methods Overview*
*Interact J Med Res 2024;13:e59772*
*URL: https://www.i-jmr.org/2024/1/e59772*
*doi:10.2196/59772*
*PMID:*

Review

# Visual Modeling Languages in Patient Pathways: Scoping Review

Binyam Bogale[1*], MPH, MSc, PhD; Märt Vesinurm[2], MSc; Paul Lillrank[2], PhD; Elisabeth Gulowsen Celius[3], MD, PhD; Ragnhild Halvorsrud[4*], PhD

[1]Department of Neurology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[2]Institute of Healthcare Engineering and Management, Department of Industrial Engineering and Management, Aalto University School of Science, Espoo, Finland

[3]Department of Neurology, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[4]Sustainable Communication Technologies, SINTEF Digital, OSLO, Norway

[*]these authors contributed equally

**Corresponding Author:**
Binyam Bogale, MPH, MSc, PhD
Department of Neurology
Institute of Clinical Medicine
University of Oslo
Bldg 20
Ludvig Kastensvei 15
Oslo, 1064
Norway
Phone: 47 96711061
Email: binyadad@gmail.com

## Abstract

**Background:** Patient pathways (PPs) are presented as a *panacea solution* to enhance health system functions. It is a complex concept that needs to be described and communicated well. Modeling plays a crucial role in promoting communication, fostering a shared understanding, and streamlining processes. Only a few existing systematic reviews have focused on modeling methods and standardized modeling languages. There remains a gap in consolidated knowledge regarding the use of diverse visual modeling languages.

**Objective:** This scoping review aimed to compile visual modeling languages used to represent PPs, including the justifications and the context in which a modeling language was adopted, adapted, combined, or developed.

**Methods:** After initial experimentation with the keywords used to describe the concepts of PPs and visual modeling languages, we developed a search strategy that was further refined and customized to the major databases identified as topically relevant. In addition, we consulted gray literature and conducted hand searches of the referenced articles. Two reviewers independently screened the articles in 2 stages using preset inclusion criteria, and a third reviewer voted on the discordance. Data charting was done using an iteratively developed form in the Covidence software. Descriptive and thematic summaries were presented following rounds of discussion to produce the final report.

**Results:** Of 1838 articles retrieved after deduplication, 22 satisfied our inclusion criteria. Clinical pathway is the most used phrase to represent the PP concept, and most papers discussed the concept without providing their operational definition. We categorized the visual modeling languages into five categories: (1) general purpose–modeling language (GPML) adopted without major extension or modification, (2) GPML used with formal extension recommendations, (3) combination of 2 or more modeling languages, (4) a developed domain-specific modeling language (DSML), and (5) ontological modeling languages. The justifications for adopting, adapting, combining, and developing visual modeling languages varied accordingly and ranged from versatility, expressiveness, tool support, and extensibility of a language to domain needs, integration, and simplification.

**Conclusions:** Various visual modeling languages were used in PP modeling, each with varying levels of abstraction and granularity. The categorization we made could aid in a better understanding of the complex combination of PP and modeling languages. Standardized GPMLs were used with or without any modifications. The rationale to propose any modification to GPMLs evolved as more evidence was presented following requirement analyses to support domain constructs. DSMLs are infrequently used due to their resource-intensive development, often initiated at a project level. The justifications provided and the context where DSMLs were created are paramount. Future studies should assess the merits and demerits of using a visual

modeling language to facilitate PP communications among stakeholders and use evaluation frameworks to identify, modify, or develop them, depending on the scope and goal of the modeling need.

***(Interact J Med Res 2024;13:e55865)*** doi:10.2196/55865

**KEYWORDS**

patient pathways; visual modeling languages; business process model and notation; BPMN; unified modeling language; UML; domain-specific modeling languages; scoping review

## Introduction

### Background

The concept of patient pathways (PPs) has been widely used to improve health system functions across a range of health conditions, care levels, and regions [1]. PP implementation spans from emergency care [2] to specialized fields such as cancer care [3,4], with recent adaptations to address conditions like the COVID-19 pandemic [5]. Effectiveness reviews of PP implementation indicated improved patient outcomes, reduced length of stays and cost of care, enhanced teamwork, and improved documentation [6-11]. Depending on the type of pathology and nature of the organization where the concept was introduced, the evidence on the effect of the outcomes is inconclusive. The confusion around the concept [12], the variability in its quality [13], and the deficiencies in the process of contextualization were among the implicated factors in the effectiveness studies that reported on patient, health system, and finance outcomes [14]. Seys et al [14] indicated the dual complexity, that is, the PP itself is a complex concept implemented in a complex health system. This calls for clarity in all aspects, including simplification of the description to facilitate communication and common understanding to enhance its effectiveness [15-17].

Multifaceted factors pose varying challenges to maximizing the benefits of PPs. Yet, there is an ongoing discussion on the unified definition and frameworks for their development, implementation, and evaluation [12]. Efforts are being made to synthesize and consolidate the various terms used and their definitions since the first identification of several alternative names is in action [18]. Furthermore, 84 different definitions along with the differing focus of pathways in the United States and the United Kingdom were identified shortly after their use [19]. One of the pioneering definitions, also adopted by the European Pathway Association, states that “A care pathway is a complex intervention for the mutual decision-making and organization of care processes for a well-defined group of patients during a well-defined period” with characterizing features [20]. To identify studies for an effectiveness review, Kinsman et al [21] proposed an operational definition, later refined by Lawal et al [22], and validated it in an emergency medicine review [23]. The use of terms evolved from merely clinical orientation of care provision (clinical pathways) to the inclusion of more stakeholders and from elements of care organization (care pathways) to the inclusion of multiple levels of care (integrated pathways or integrated care pathways) to develop a patient-focused systemic approach (PPs). Schrijvers et al [24] argued that adding qualifiers such as in “integrated care pathway” is unnecessary because the care pathways are integrated by definition. Such comments seem not to have held on because the most recent concept analysis paper proposed an even longer term, a “patient-centered care pathway,” showing an increase introduction of terminologies [12]. Their proposed definition, “a long-term and complex managerial intervention adopting a systemic approach, for a well-defined group of patients who journey across the entire continuum of care, from prevention and screening to recovery or palliative care,” with several attributes [12], however, is indicative of perspectives added toward comprehensive and patient-centric concepts. The key characteristics and elements were listed, and their importance was stressed beyond the proposed definitions in each article [12,19,21,22], with slight variations. In this review, we use “patient pathways” to represent the concept from a patient-centered care perspective [12] while using the criteria proposed for a Cochrane systematic review [21,22] to identify the articles. Patient journey studies, which focus on patient-centric mapping and analysis of health care delivery processes, are increasingly being introduced to the scene [25]. To simplify, we use “patient pathways” to represent a plan as a blueprint of the care process, while patient journey denotes an individual experience of the planned PPs revealed retrospectively. In this review, we used different terms interchangeably in the identification and review of articles. Confusion in the definition and conceptualization leads to variabilities in the analysis and modeling of PPs [12].

The rampant siloed and local productions with varying representations of a PP [1] call for more standardized ways to describe and communicate PP to have a shared understanding of the concept. This can be argued in the same manner as consensus frameworks have been proposed [12,14,26,27] in an attempt to standardize the stepwise development, implementation, and evaluation of PPs. The standardization process starts with the modeling languages that are used to describe and communicate PPs. Particularly, visual modeling “not only enables easy interpretation but moreover denotes a useful means for communication and understanding” [28] of a process. Graph-based formalism is one of the two most common process-modeling approaches [29]. The benefits of visual modeling over other forms of representation have been extensively discussed in business process modeling, although there remains a scarcity, particularly in the PP domains, of effectiveness studies using a traditional approach [29-31]. Lack of a common language also exacerbates the interoperability challenge in the increasingly digitized care processes, including digitized PPs integration to other electronic patient records and quality improvement digital tools [32,33].

## Prior Systematic Reviews on Modeling Languages

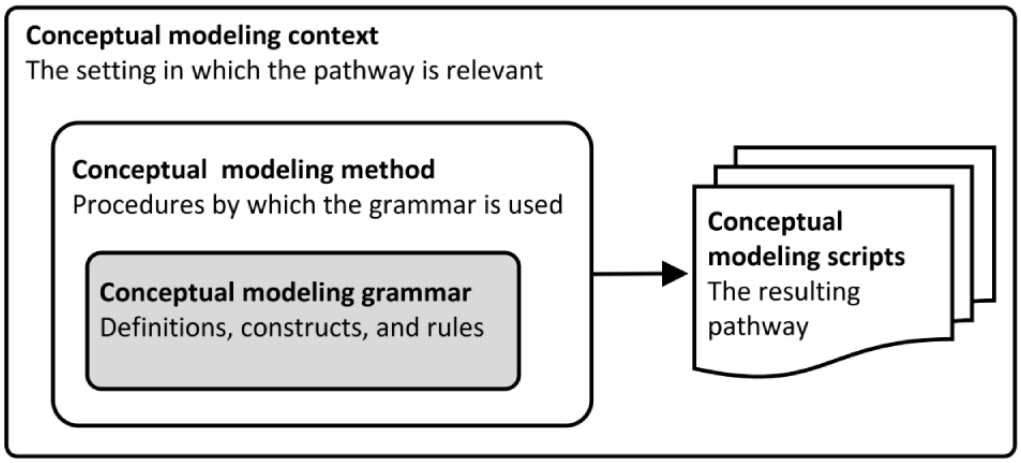
The modeling of PPs can be done to understand and analyze the current state, which is often referred to as an "As-Is" model, or a redesigned or improved PP modeling can be done in a "To-Be" model. The "As-Is" model is often "data-driven" modeling that uses clinical data from various electronic tools, for example, electronic health records (EHRs) and registers. Clinical data-based mathematical modeling often focuses on exploring the patient journey retrospectively, the focus being on simulation studies, data mining, machine learning, and artificial intelligence to predict the most efficient path for a care process [33,34]. The "To-Be" model considers PPs as a complex concept, which is prescriptive in nature and created by a multidisciplinary team. One of the challenges for this model is the absence of a common process-modeling language that is complex enough to incorporate all the necessary aspects in a model while being simple enough to be used by nonmodeling domain experts, either manually or digitally [35].

There are many different process-modeling languages coming from different scientific traditions, and their characteristics and intended use vary greatly [36-38]. However, the complex nature of health care delivery has led to a modest uptake of business process–modeling languages that are shown to be useful in sectors with predictable work processes [39]. Although choosing a modeling language has generally been a challenge [37,38], there are a few systematic reviews that have targeted specific modeling languages or notations in the care process [39-41]. Specific to PP-related concepts, there is a comprehensive review on modeling [42] while other reviews are more inclined toward data-driven pathway description and optimization [36]. The most similar review to our scoping review was conducted by Mincarone et al [37] in which only standardized modeling languages were included on the entirety of care process, which is wider in scope than PPs. To improve the domain expressiveness of a modeling language, extending or combining a general purpose–modeling language (GPML) or developing a domain-specific modeling language (DSML) can often be done. To our knowledge, the justification for and the "how" of such efforts have not previously been summarized. In addition, focusing on standardized languages and notations potentially excludes DSMLs, which are costly to develop but may display higher expressiveness than the GPMLs, including their domain-extended versions.

## Goal of This Scoping Review

This review's scope is to include both DSMLs and standardized GPMLs with emphasis on how they were used to meet the domain-specific requirements. Our focus is on visual modeling languages but not on 1D textual languages as delineated by Moody [43]. Visual languages in this case include ontological modeling languages that depict terms and concepts with their relationships in a visual manner. Most of the PPs are described in natural languages, with flowcharts and tables accompanying for simplicity, but we did not include simple flowcharts and diagrams as a visual notation. Wand and Weber [44] introduced a framework for research on conceptual modeling consisting of 4 elements to ease communication. To reduce potential confusion around terms in this scoping review, we introduce an adapted version of this model (Figure 1, adapted from Wand and Weber [44]). The modeling *grammar* constitutes the inner core because it provides definitions, constructs, and rules to produce a model. The modeling *script* is the end product of the modeling process. The modeling *method* describes how grammar can be used to produce a script. The modeling *context* has a wider perspective and describes the setting in which the modeling occurs. We do not intend to provide a comprehensive review of the *script* because our focus lies at the core of the framework. While grammar is the core, modeling method and context are of interest in this review because this review aims to contribute to the different ways in which a modeling language is presented.

**Figure 1.** Framework for research on conceptual modeling (adapted from Wand and Weber [44]).




## Review Questions

A scoping review was conducted to systematically identify and map visual modeling languages used to describe and communicate PPs. The review addresses the following review questions:

- Review question 1: Which visual modeling languages were used in the modeling of PPs?
- Review question 2: What are the justifications provided to adopt, adapt, or develop a visual modeling language to describe and communicate PPs?

- Review question 3: Within what contexts are the visual modeling languages applied in PP projects?

## Methods

### Overview

The reporting of this review follows the systematic review extension for scoping reviews PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [45]. We followed the methodological framework proposed by Arksey and O'Malley [46], which was further improved by Levac et al [47] and the Joanna Briggs Institute [48], to conduct this scoping review. The review questions were proposed and refined after a thorough exploration of the existing synthesis on the topic, as presented earlier. The rest of the framework is integrated into the reporting format as presented in subsequent sections.

### Eligibility Criteria

To be included in the review, papers needed to satisfy the operational definitions (Textbox 1) for both PP and a visual modeling language. The article's main aim could be either a theoretical discussion or empirical research on a modeling language that is created, adopted, adapted, or combined. Peer-reviewed journals, conference and congress proceedings, and gray literature that are accessible on the internet, regardless of the geographical location, type of health condition, and publication date up until the last systematic search dated November 21, 2022, were eligible. For papers published in a conference proceeding, we searched for and opted to include the more complete peer-reviewed publication in a scientific journal, if it existed, in expectations of detailed information. Only papers published in English were considered. No restrictions on study design, population characteristics, type of health care facility, or level of care were applied.

Papers were excluded if their focus was mainly on mathematical models, simulation studies, or machine learning without having information to report on the visual graphical presentation of a model for facilitating communications, including to nonmodeling expert stakeholders. Comprehensive reviews [35,36,42] have been done to cover modeling methods and languages applied in data-driven retrospective models. According to the conceptual framework proposed by Wand and Weber [44], we excluded papers that have primarily emphasized the presentation of a pathway using a known modeling language without an adequate description of how it is being used differently in the domain it is being applied to. Papers uniquely reporting modeling tools (software artifacts) without discussing the underlying modeling language used, or its semantics and ontologies for reasons of extension or visual presentation were excluded. Ontology-based and semantic modeling languages, with or without the inclusion of visual modeling as an output, were included in the review given the aim of this study is to facilitate communication of PPs among the stakeholders. According to a review by Zarour et al [49], representation of business process model and notation (BPMN) extensions can be of one of the 3 formats: metamodel, XML Schema, or graphical elements, and therefore, we included metamodels in this review.

We excluded simple flowchart presentation of pathways, which is the most common description and communication approach in medical domain but is now headed toward extinction [42] because the more mature notations were extensions of it and the relevance of including and discussing its use was perceived to be not adding to the standardization and wide acceptance of visual modeling for PP discussion. The enforcement of inclusion criteria began with the development of search strategies broadly and gradually narrowed down during the selection process. In the initial screening stage, articles were included if they mentioned or described 2 concepts, PPs and visual modeling language, in the abstract. The second stage involved a closer examination of these concepts. For instance, papers mentioning terms related to PPs but discussing broader concepts such as general care processes were excluded. Similarly, articles mentioning a visual modeling language but failing to describe how it was used in their research context were also excluded.

**Textbox 1.** Components and Definitions according to population, concept, and context framework proposed by the Joanna Briggs Institute.

**Population and Participants**

- Regardless of the professional background of the experts behind the modeling language; irrespective of the health condition and level of care. The paper can be a theoretical exploration of how existing process-modeling languages, particularly general purpose–modeling languages (GPMLs) applied to other domains, can be adopted or adapted. This extends to the development of domain-specific modeling languages (DSMLs) for describing and communicating patient pathways (PPs) to facilitate common understanding among intended stakeholders.

**Concept**

Two concepts are as follows:

- Patient pathways: We based the PPs concept definition on a refined operationalization criterium by Kinsman et al [21] and by Lawal et al [22] in that they used it to identify articles for a Cochrane systematic review. Articles primarily focusing on theoretical discussions of a modeling language without empirical studies (including a case study for the sake of demonstration) might not reflect the aforementioned criteria. Such articles were included as long as the authors clearly stated that the application is for patient pathway concepts. As the terms used have evolved through time and are sometimes used interchangeably, we relied on the concept definition and explanations by the authors. Papers containing terms and phrases that are often used interchangeably, such as care process, workflow, etc were excluded if the paper did not explicitly state the PP concept dealing both with the organization and clinical part of the care.
- Visual modeling language: general-purpose or domain-specific modeling languages that may or may not be standardized, can be graphical, rule-based, or combined presentations, aimed at describing a PP, regardless of the origin and the extent to which a given modeling language has been implemented. The modeling language can be adopted, adapted, combined (as in complementary combinations of stand-alone modeling languages), or developed to model a PP and can or cannot have digital applications and tools accompanying the language. Modeling languages used in mathematical models, simulations, artificial intelligence, and machine learning are outside the scope of this review

**Context**

- The modeling language applied regardless of disease or health condition, treatment and intervention options, clinical settings, and service delivery level (primary, secondary, or tertiary care facilities). No restrictions based on geographical location and scope within the location.

## Information Sources

We conducted searches in the following databases: MEDLINE via PubMed, PsycINFO, Embase, CINAHL, and Scopus after iteratively developing search strategies. We used Joanna Briggs Institute [48] population, concept, and context framework to exhaustively list search terms under each component. The following initial search terms were used: “Patient Pathway*,” “Care Pathway*, “Clinical pathway*,” along with the Medical Subject Heading term “Critical pathway/” in MEDLINE and CINAHL databases to decide on what additional terms to combine with them to identify papers relevant to our inclusion criteria. For words describing modeling language and related concepts, terms like “language*,” “model*,” “framework*,” “formalism*” were used and search was expanded by including additional similar terms from retrieved articles. We narrowed the scope of the terms by adding descriptors. The search strategies were drafted in consultation with an experienced librarian and customized to each database after iterative rounds of improvements. We limited our search strategy to the title and authors’ keywords after discovering that including the abstract greatly increased the number of search hits. The final search strategies for each selected database are available in Multimedia Appendix 1. Gray literature search and hand searches of relevant articles referenced by included papers were done by the first author. After learning that the IEEE website has useful collections on the topic, we made searches in addition to the initial database search. As a forward search strategy, Web of Science and Google Scholar databases were used.

## Selection of Sources of Evidence

The librarian used EndNote (Clarivate) to deduplicate the retrieved articles. After importing them into Covidence (Veritas Health Innovation Ltd) software, any remaining duplicates were identified and automatically removed using the software’s built-in feature. In total, 2 reviewers independently screened publications in 2 stages after importing all the retrieved papers into the Covidence software [50]. To have consistency in the screening process, we defined and discussed the inclusion and exclusion criteria and documented them in the Covidence software for easy referencing and to use the automated word detection features of the software. Accordingly, the 2 reviewers conducted the screening of the title, index words, and abstract at stage 1 and full-text review screening at stage 2 guided by the predetermined inclusion and exclusion criteria. In case of discordance, a third reviewer voted the article as in or out in stage 1 screening, and a consensus-based resolution of disagreements was reached in stage 2 between the 2 reviewers.

## Data-Charting Process

A data-charting form was developed by the first author and discussed with the second reviewer who participated in all the screening stages. The 2 reviewers discussed and iteratively improved the data-charting form while retrieving the results from the included papers. The data-charting form was developed to capture key information about the modeling language in addition to the identifying characteristics of the papers (author, title, date of publication, region or country, and journal or source). The form featured a free text section dedicated to capturing details about the modeling language including its name, descriptions, justification, and its application within the research context. In addition, categorical information, such as yes or no responses, was included accompanied by a follow-up free text area. For example, questions like “Is there an associated digital tool?” allowed for documentation of tool names, if applicable. Regarding the concept of PPs, we extracted the

phrases used to define or represent the concept, as well as any stated definitions within the document. All the data-charting process was conducted using the Covidence software and eventually exported to a Microsoft Excel spreadsheet.

### Synthesis of Results

The synthesis of the results followed the review questions where we descriptively categorized the modeling languages into thematic groups. We also thematically analyzed the justification, purposes, and context in which the modeling language was created or implemented.

## Results

### Selected Articles

We imported a total of 1835 articles from the selected databases to Covidence after removing duplicates (Figure 2). In addition, we discovered 5 more articles through hand searching, including gray literature searches from the Object Management Group (OMG). Of the 22 papers included in our review, half were published in conference proceedings and mainly retrieved from IEEE. The other half, except for 1 article from gray literature, were obtained from peer-reviewed scientific journals (Table 1). We included papers that were published as far back as 2008 to as recent as 2022. Most of the papers were from Europe (Germany and Italy), and almost all of them were from high-income countries. We found that more than 1 modeling language was contributed by similar groups of coauthors in another paper included in the review.

**Figure 2.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

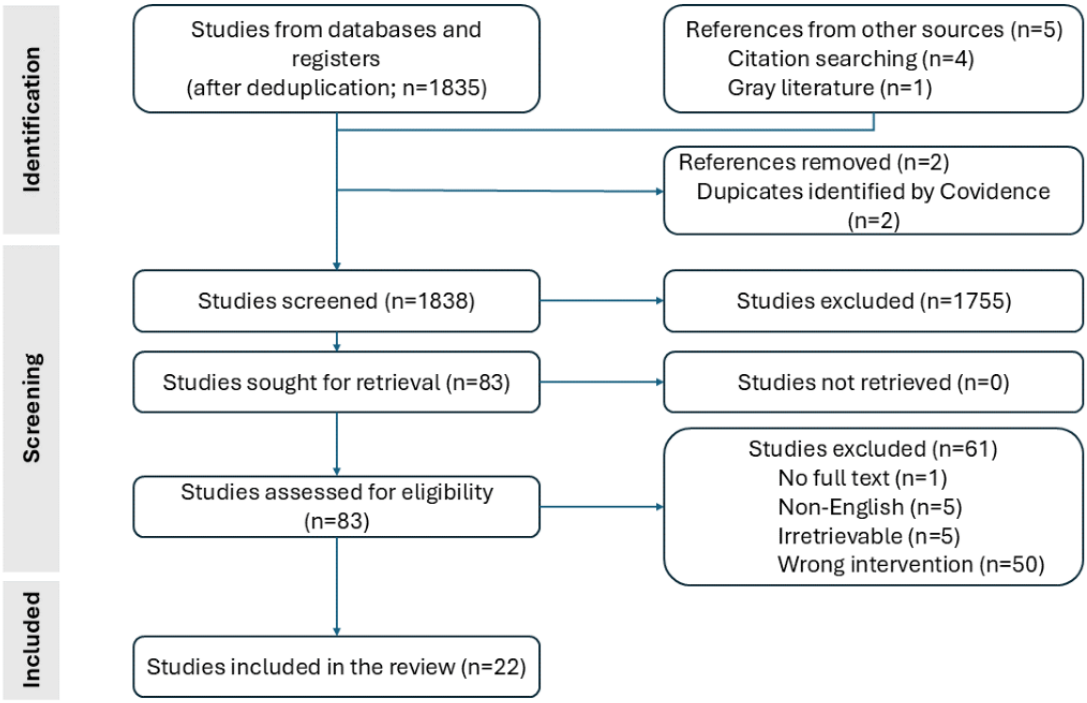


**Table 1.** Summary information of the included publications.

| Study | Country | Publication source (categorized) | Modeling language used | Case study or demonstrated on |
|---|---|---|---|---|
| Scheuerlein et al [51], 2012 | Germany | Journal | *BPMN*[a] (*tBPM*[b]) | Colon and rectum carcinoma (treatment of) |
| Barbagallo et al [52], 2015 | Italy | Journal | *BPMN* | Multiple surgical conditions |
| Ferrante et al [53], 2013 | Italy | Journal | *UML* [c] | (Post) stroke rehabilitation |
| Ferrante et al [54], 2016 | Italy | Journal | *UML* | Stroke rehabilitation |
| Mauro et al [55], 2010 | Germany | Proceeding[d] | *UML* | Not mentioned |
| Zerbato et al [56], 2015 | Spain and United States | Proceeding | *BPMN for process and UML for data* | Catheter-related bloodstream infections |
| Braun et al [57], 2016 | Germany | Proceeding | *BPMN4CP 2.0* [e] | Stroke |
| Richter and Schlieter [58], 2019 | Germany | Proceeding | *BPMN, with Quality BPMN, with quality indicator extension* | Integrated stroke care |
| Tehrani et al [59], 2012 | United Kingdom | Proceeding | *BPMN, with norm extension* | Major gynecological surgery |
| Combi et al [60], 2017 | Italy | Journal | *BPMN and DMN* [f] | COPD[g] |
| Sooter et al [61], 2019 | United States | Journal | *BPMN and DMN* | Contraceptive use |
| Object Management Group Healthcare Domain Taskforce [62], 2020 | United States | Gray | *BPM+health* [h] | Multiple conditions |
| Bowles et al [63], 2018 | Germany | Proceeding | *BPMN and LES* [i] | Diabetes and hypertension |
| Ardito et al [64], 2020 | Italy | Journal | *Metamodel; EER* [j] *, BPMN and UML* | Headaches |
| Iglesias et al [65], 2022 | Spain | Journal | *TP-VML* [k] | Catheter-related bloodstream infection |
| Trajano et al [66], 2021 | Brazil | Journal | *MedPath* | 85 different care pathways: including low-back pain, diabetes, syphilis, etc |
| Burwitz et al [67], 2013 | Germany | Proceeding | *CP-Mod* | Tooth extraction |
| Shitkova et al [68], 2015 | Germany | Proceeding | *Icebricks* | Cardiac diseases |
| Li et al [69], 2008 | China | Proceeding | *OWL* [l] | None, theoretical |
| Ye et al [70], 2008 | China | Proceeding | *CPO* [m] *, domain ontology, and SWRL* [n] | None, theoretical |
| Nishimura et al [71], 2014 | Japan | Proceeding | *CHARM* [o] *: ontological* | 7 different diseases |
| Abidi and Abidi [72], 2012 | Canada | Proceeding | *OWL* | Prostate cancer |

[a]BPMN: business process model and notation.

[b]tBPM: tangible business process modeling.

[c]UML: unified modeling language.

[d]Includes congress, conference, symposium, and published papers mainly on the IEEE and IEEE Xplore websites.

[e]BPMN4CP 2.0: business process model and notation for clinical pathways 2.0.

[f]DMN: decision model and notation.

[g]COPD: chronic obstructive pulmonary disorder.

[h]BPM+Health: business process management for health care.

[i]LES: labeled event structure.

[j]EER: enhanced entity relationship.

[k]TP-VML: task planning visual modeling language.

[l]OWL: web ontology language.

[m]CPO: clinical pathway ontology.

[n]SWRL: semantic web ontology rule language.

[o]CHARM: convincing human action rationalized model.

## Modeling Languages

### Overview

Modeling languages that have their origin and wider uptake outside of the health domain and others developed within health domains were identified. We have collectively called the former GPML, and the latter DSML. In this context, GPML means a widely used modeling language or notation to visualize processes that transcend the health domain. These modeling languages can be used in the health care domain either (1) directly and without major extension or formal combination, (2) by extending the specifications following extension recommendations, or (3) by combining to enhance their expressiveness. On the basis of the presentation of the output, some can be formal notations and others visually presented concept relationships. To provide a detailed and simplified account, we grouped the included modeling languages into categories mentioned in subsequent sections.

### GPMLs: Without a Formal Extension to the Domain Requirement

This category contains adopted process-modeling languages that are standardized and widely applied in other sectors. The 2 OMG standards, BPMN [51,56] and unified modeling language (UML) [53-55], were used without any demonstrated extension or modification to the original specifications. The methodology to enhance GPML uptake, especially to include nonmodeling experts in the modeling process, was not considered as a change to the modeling grammar and syntax. In the case of tangible business process management (BPM) [51], the authors used physical icons instead of digital tools that enhanced the participation of the domain experts. Zerbato et al [56] stated extending with additional time primitives to represent temporal constraints, which is not supported by the BPMN version that they have used. However, the details of the extension and its outcome were not provided. In the latest of the 2 papers [54], Ferrante et al [53] discussed the method around the modeling process to the original paper where UML is used to model the stroke rehabilitation pathway.

### GPMLs: With a Formal Extension to the Domain Requirement

By adapting the widely used standardized notation, Braun et al [57] provided an extension formalism called BPMN4CP (business process model and notation for care pathways) to include care pathway—specific aspects. The revised BPMN4CP 2.0 covers additional domain-specific requirements to the original recommendations. Whereas Richter and Schlieter [58] extended BPMN to add the quality indicator specifications based on the BPMN4CP extension framework. Tehrani et al [59] extended the BPMN specification from the results of their norm analysis, together with organizational semiotic methods, which focus on describing the human behavior and conditions under which the human behaviors occur. In the notation, the extension was indicated by labeling (N#). In all 3 cases, it is imperative to take advantage of the widely used process models with their extension possibilities.

### A Combination of Modeling Languages

To satisfy the domain requirements and specific needs that arise in the modeling processes, 2 or more visual modeling languages were used in combination, often complementing one another as required by the nature of each part in a given model. The OMG BPM+health provides a possibility to combine 3 independent modeling languages (BPMN, case management model and notation [CMMN], and decision model and notation [DMN]) [62], while Combi et al [60] and Sooter et al [61] combined BPMN and DMN in their respective studies. Extensive use of other modeling languages in combination, including the use of data modeling to identify comorbidity was presented in 2 instances [63,64]. One could argue that these may not be identified as visual modeling languages.

### Developed DSMLs

Three articles [65-67] discussed the development of a DSML; one was called MedPath, which was specifically developed for the PP modeling process [66], and the other 2 radically improved [65,67] the base modeling language that inspired the development of specific DSML to model PPs.

### Ontological Modeling Languages

This category is made acknowledging the differences between the promises of visual modeling and ontologies. While visual models do not intend to present a complete description of the domain and the constructs, rules should reflect some ontological commitments. Ontological models go beyond such restriction and cover a possible set of concepts and premises in a domain [72]. Ontology-based modeling languages were supported by other methodological approaches, such as semantic and norm analyses, and specifically dealt with terms and expressions to satisfy the PP requirements. Li et al [69] built an ontology chart that is presented in a diagrammatic visualization of the constructs following semantic and norm analysis. The widely used web ontology language and Protégé (a free, open-source ontology editor and framework for building intelligent systems) were used to personalize the pathways. To align institution-specific PPs in an automated solution, Abidi and Abidi [72] used semantic web-based modeling using web ontology language to align pathway ontologies. In another study, semantic web ontology rule language was used to represent the temporal relationship that was not covered by the time subontology of the clinical pathway ontology together with domain ontology [70]. Aimed at confirming the practical ability of convincing human action rationalized model to represent medical actions, promoting knowledge sharing, and inheritance in a computer-interpretable way, Nishimura et al [71] built the convincing human action rationalized model tree that contributed to finding the commonalities, variations, and reasons for differences among pathways.

## Justification to Adopt, Adapt, Combine, or Develop Modeling Languages for PPs

The justification provided by authors for the identification, selection, use, and development of each modeling language varied across many different factors. These include but are not limited to the aim, level of analyses, the composition of the group behind the language, and when it was proposed during the progressive maturity of visual process languages.

### *Justification to Adopt a Visual Modeling Language*

The justification given to adopt the GPMLs without extension was mainly to test the applicability of the modeling languages in health care domain and how the digitization process can be supported. The popularity and wider tool support made it easier to identify from the vast selection that already exists. Some authors provided comparative analysis to justify their choice.

### *Justification to Extend a Visual Modeling Language*

To extend a modeling language, the main rationale emanates from the understanding of the deficiencies of GPMLs to meet the PP domain requirements. Apart from the popularity and resources around the modeling language, the presence of extension metamodel and the frameworks to guide the extension process are necessary to begin assessing whether the requirements can be met without extension (Table 2).

**Table 2.** The justification to extend a modeling language.

| Study | Modeling language | Justification to adapt or extend |
|---|---|---|
| Braun et al [57], 2016 | BPMN4CP[a] | • Domain requirements are not fully represented, and the extension procedure needs to follow a framework |
| Richter and Schlieter [58], 2019 | BPMN[b], with quality indicator extension | • Widely accepted and established standard<br>• Gives a metamodel, suitable for extension<br>• Presence of extension framework [56] to build on |
| Tehrani et al [59], 2013 | BPMN, with norm extension | • BPMN is “a rigorous method that provides a rich set of techniques and notations for process modeling”<br>• Absence of techniques to describe human behavior and the conditions under which the behavior occurs |

[a]BPMN4CP: business process model and notation for clinical pathway.

[b]BPMN: business process model and notation.

### *Justification to Combine Visual Modeling Languages*

No single modeling language can adequately cover the requirements because the health care domain in general, and the PP modeling in particular, is a complex process. In this category (Table 3), the main rationale for combining modeling languages emerges from the conviction that complementing the deficiency of one model with the strength of the other is possible. In three of the cases, all the combinations were made progressively and included the specifications from OMG [60-62].

The health care Domain taskforce of the OMG has introduced a field guide called *shareable clinical pathways version 2.0* introducing amalgamation of their 3 standard notations for clinical pathway modeling [62]. Although not officially endorsed as a standard, it serves as a valuable discussion paper. The primary objective is to establish a modeling technique that is universally comprehensible among various stakeholders, including business analysts, health care professionals, and IT developers.

Implementing health information technology based on a care model that is universally developed and understood contributes to the efficiency, cost-effectiveness, and quality of health care delivery. The guide proposes the use of BPMN for prescribed models, CMMN as a complement for actual workflows, and DMN to address decision modeling, encompassing a complex set of factors to arrive at the most appropriate clinical decision. The suggested combination facilitates the creation of business flow diagrams and decision tables, offering coverage for clinical, administrative, and revenue cycle processes. This model, using a blend of these languages, addresses diverse aspects essential in the health care domain and is standardized for seamless sharing across organizations.

The BPM+ community of interest was assigned the task of examining the alignment between 2 OMG standard notations, namely BPMN and DMN. The focus was on bridging the gap between narrative guidelines and their digital representation with the aim of achieving a “true integration” of guidelines and pathways within electronic medical record systems. Sooter et al [61] illustrated the feasibility of modeling a clinical guideline for contraception in a standardized format suitable for digitization. Their approach involved listing and defining all data points using a spreadsheet, which subsequently indicated an area for extension to the BPMN 2.0 specification. The team showcased the challenges encountered in mapping SNOMED-CT (Systematized Nomenclature of Medicine–Clinical Terms) codes and presenting all terms in a summary chart. The decision logic and the business process aspects related to contraceptive choices and delivery were effectively modeled with a low level of abstraction using a modeling tool.

Combi et al [60] developed a framework using DMN to model decision-intensive care aligned with clinical practice guidelines directed at clinicians. Simultaneously, they used BPMN for modeling the care organization aspect. The authors rationalized the misuse of BPMN for modeling the entire pathway, including

clinical decisions, to conduct their research. However, they acknowledged the limitations of BPMN, particularly its shortcomings in fully supporting temporal relationships, domain knowledge, and the integration of complex structural data. The discussion explored the potential use of extensions to address these drawbacks. The steps outlined in the framework are more suitable for stakeholders with modeling expertise because the implementation of the modeling tool demands such skills.

Ardito et al [64] argued that finding a modeling language that is detailed enough to balance a machine-executable model yet simple enough to be human understandable is a challenging task. They proposed a modular approach for executing complex machine-level processes separately that uses a task-oriented chatbot approach based on the modeled pathways while interacting with the social media chatbot with the end user. The process used a stepwise approach and is method-intensive by nature, and it also includes patients, who are often forgotten as stakeholders in modeling efforts.

**Table 3.** The justification to combine modeling languages.

| Study | Modeling languages | Justification to combine |
|---|---|---|
| Combi et al [60], 2017 | BPMN[a] and DMN[b] | • Merely using a BPMN to model a decision-intensive care pathway is a misuse of the specification<br>• Need for a framework that guides the use of DMN for a decision and BPMN for a structured process |
| Sooter et al [61], 2019 | BPMN and DMN | • A “true integration” with electronic health records is not yet achieved<br>• The compatibility of the 2 modeling languages to change narrative guidelines to digital instantiations is not checked |
| Object Management Group Healthcare Domain Taskforce [62], 2019 | BPM+Health[c] (BPMN, CMMN[d], and DMN) | • Health care domain needs process models for prescriptive workflows, case models for reactive workflows, and decision models for complex business rules, hence appropriate modeling languages<br>• Need to use accepted standards to make pathways shareable |
| Bowles et al [63], 2017 | BPMN and LES[e] | • Other available presentations of CPsf for multimorbidity have inherent ambiguity. There is a need to resolve pathway conflicts using standardized, coordinated back-end and front-end models |
| Ardito et al [64], 2020 | Metamodel; EER[g], BPMN, and UML[h] | • Finding a balance between modeling language expressiveness and the automated execution of modeled processes is difficult, and investing in it is unprofitable. To find the balance that does not require the domain experts to adopt a complex process-modeling language, what about dedicating the burden to an executing module—a chatbot engine? |

[a]BPMN: business process–modeling notation.

[b]DMN: decision model and notation.

[c]BPM+Health: business process management for health care.

[d]CMMN: case management model and notation.

[e]LES: labeled event structure.

[f]CP: clinical pathway.

[g]EER: enhanced entity relationship.

[h]UML: unified modeling language.

### *Justification to Develop a DSML*

Justifications to develop a modeling language included verbosity and issues related to customization and execution, limitations to cover PP domain requirements sufficiently, and needs such as integration. Furthermore, using domain-oriented existing languages was mentioned in this context (Table 4).

MedPath [66] is a process-based DSML that was developed to capture all the components in a clinical context while minimizing the verbosity challenges encountered by adopting the GPMLs. The language has syntax, semantics, and a visual notation. MedPath is a layer between the expert who develops a model and the engine that translates the metamodel into visual elements to be easily understood by the domain experts. The authors stated that the language is designed to be comprehensive enough to capture all aspects of care organization, simple enough to be understood by health care domain experts and detailed enough to be integrated into the health information system.

The *openEHR* foundation strives for interoperable EHR systems and the use of standardized models in care processes and is also behind the task planning (TP) initiative. TP is a clinically oriented specification that also has a visual modeling language called task planning visual modeling language (TP-VML) [65]. The authors credit the BPMN extensions that are aimed at solving the temporal constraints to model clinical workflow. However, TP formalism can better represent the health care domain because it has a more domain-specific orientation than the extensions. The original TP-VML icons and semantics are not discussed, while the displayed visual notation in the paper

references its source to the tool that the authors used. Possible extensions of the TP specifications were presented after a thorough analysis following the stepwise methods. The authors concluded that following the extension procedures suggested for BPMN to pathways by Braun et al [57], TP can be extended to include domain requirements and can be used for complex PPs as demonstrated in the case study on the catheter-related bloodstream infection. Future work should focus on furthering the decision logic specification to evoke rules from TP supported by an expression language and a basic metamodel of openEHR.

Burwitz et al [67] developed a DSML named CP-Mod based on the clinical algorithms basic process-modeling concepts and extended to include evidence-based medicine and decision support, classification of treatment alternatives, and time events and waiting periods following the requirement they set for PP modeling. The justification for using the base modeling concept and the need for extension in comparison to the widely used modeling languages were presented, and good arguments were made.

Icebricks [68] is a modeling methodology and digital tool that also provides a human-readable representation of care pathways. It meets the need for a notation that depicts activities, annotates information, and supports standardization while facilitating collaborative work with easy learning. The model can be exported to Microsoft Word with full process documentation, including diagrams (although not shown in the paper) and annotated information. The authors raised important questions that qualify a useful notation, but more emphasis was given to the methodology and the tool than to the core of the language.

**Table 4.** The justification to develop a modeling language.

| Study | Modeling language | Justification to develop |
|---|---|---|
| Iglesias et al [65], 2022 | TP-VML[a] | • More clinically oriented business process management standards can provide an adequate representation of the temporal orientation of clinical workflow than BPMNb or its extension. |
| Trajano et al [66], 2021 | MedPath | • GPMLsc verbosity and not easily customized to medical context.<br>• Also sometimes lack infrastructure for integration and execution. |
| Burwitz et al [67], 2013 | CP-Mod | • Reviewed common modeling languages used against their set requirement and identified deficits.<br>• Clinical algorithm can be used as a base and organizational aspect and individual data can be added. |
| Shitkova et al [68], 2015 | Icebricks | • Generic modeling languages did not sufficiently cover PPd requirements.<br>• A notation needs to allow the representation of activities and process flows, annotate relevant information, and represent knowledge on an appropriate level of abstraction. |

[a]TP-VML: task planning visual modeling language.

[b]BPMN: business process model and notation.

[c]GPML: general-purpose modeling language.

[d]PP: patient pathway.

## The Contexts in Which the Languages Were Applied

The context in which the modeling process occurred can take several aspects. Here, we focus on the profile of involved experts, the organizational structure, the coverage and scope, the tool support, the nature of the study, and the level of standardization, among other relevant information to a language. Additional discussion of the context is provided in Multimedia Appendix 2 [57-68] and for selected papers in the earlier section.

Almost all articles were method-intensive and primarily focused on theoretical discussion with empirical demonstration on selected medical conditions. Only 3 articles discussed the topic without demonstrating on a case [55,69,70]. While the majority demonstrated on a single condition, a few attempted the application on several conditions [52,63,66,68,71]. The field guide [62] and contraceptive guideline modeling [61] are meant to be applied widely at several levels of abstraction. One paper focused on interinstitutional care standardization [72], integration of PPs to deal with comorbidity [63], and that it can be applied at any level of abstraction.

The drivers of the modeling language creation or use are mostly modeling experts who attempt to involve novice modelers. The domain experts were also involved at some level in the presented concept or script ranging from design to the evaluation phases. Different techniques, such as making the vocabularies tangible [51] and producing a chatbot that facilitated the participation of novice modelers [64], were identified. The intention to simplify the modeling process for stakeholders with less modeling expertise was indicated by several authors. The use of modeling languages was tied to digitization in several of the cases, particularly to facilitate machine executability [52,64-66,68].

The OMG is responsible for UML, BPMN, as well as CMMN and DMN proposed for combination with BPMN [51-56,60-62]. Languages like TP-VML that have domain-specific engagement are linked to a foundation with a potential to be widely accepted [65].

Tool support is one of the main criteria for increasing the adoption and subsequent improvement of modeling languages. More than two-thirds of the modeling languages were accompanied by a tool or software artifact

[52,54,55,61-67,69-71] and the remaining did not specify in the report.

### PPs: Terminologies and Concept Definition

The terms and operationalization of the concept synonymous with PPs varied among the included articles. The most frequently used phrase was “clinical pathways” alone or interchangeably with other phrases including “care pathways” or “integrated care pathways.” There were no justifications provided as to why a given terminology was adopted. The phrase “patient pathways” appeared in only 1 paper [58]. In most of the papers, the concept of PPs (or their synonyms) was briefly discussed in the Introduction section. Almost all referenced the definitions forwarded by de Bleser et al [19], Vanhaecht [20], or Kinsman et al [21], among others. Almost all included papers lacked operationalization of the concept in relation to the context.

## Discussion

### Principal Findings

We identified and categorized visual modeling languages used in the representation of PPs. In addition to the direct adoption of standardized and widely used modeling languages originating outside of the health domain, there are extensions and combinations of languages proposed to meet the domain requirements. We identified DSMLs that are also important contributions to the discussion of PP modeling efforts. The justifications for selecting a visual modeling language varied depending on the modeling scope and goal. The rationale to propose any modification to the language evolved as more evidence was presented following requirement analyses to support domain constructs. The direct use of standardized modeling languages without any domain-specific adaptation was done mostly to test whether the standards can also be applied to the health domain. After having evidence regarding the deficiencies to fully represent the domain needs, extension by addition and combinations of more than 1 standardized visual modeling language were introduced. The presence of extension formalisms for widely used standard modeling languages, specifically BPMN, is promising, while the process of standardizing those extensions remains unclear. Standing on the shoulders of previous extensions would contribute to subsequently expanding the specification to the domain requirements with minimum effort. There are valid reasons put forward by the DSML developers, but these need further discussion considering the cost, rate of uptake, and likelihood of standardization. Given that there are more mature languages with already advanced tools, including the advantage of execution languages to automate the model, it is essential to conduct an exhaustive comparison of which languages to choose or whether there is a need to develop a new one before embarking on the long journey.

The taxonomy that we created corresponds to the approaches used to find the best possible ways to appropriately model PP. Similar approaches were used in other related reviews mainly to facilitate the ease of understanding of such a complex field [12,42]. The purpose of using a visual modeling language, otherwise stated as “the dependent variable/the design goal” by Moody [43], was reflected in the included papers in various ways. It has been used to facilitate communication between designers and domain experts [51,66], involve patients to interact with their PPs [68], facilitate the digitization of the PPs [61-67], and facilitate integration with the EHR system [61]. The reviewed papers attempted to fulfill these purposes fully or partly.

The application of standardized modeling languages in the health care domain has previously been reported in systematic reviews [38,39]. This scoping review adds *the process* in which the GPMLs were being used to model PP-specific requirements. In this review, BPMN is categorized into GPMLs because its syntax can be used in various domains despite its business process–specific nature [56]. We believe that it is a good example to illustrate how GPMLs were being used to model PPs. From early experimentations to check for its suitability to model PPs [51,52,56], followed by the identification of domain-specific requirements resulting in extension approaches [57-59], to the recent recommendation to combine other specifications to overcome its inherent limitation [60-62], BPMN presented itself as a leading standardized process-modeling language for PP. Gartner et al [12] also reported in their review that the process nature is one of the most common attributes of care pathways and is modeled using BPMN or improved by combining with decision support modeling languages. The popularity, expressiveness and extensibility, and tool support with an execution language were presented as justifications for its use, which are parts of the parameters in most evaluation criterion frameworks [28,44,73]. According to a comprehensive systematic review, the relative simplicity of use with strong research and implementation experiences in nonhealth domains supported by OMG is convincing to agree to the notion that one should justify why they did not use BPMN over their chosen modeling language [74]. However, to fully represent the domain requirements, combinations of standards that cover aspects other than predefined and stable processes must be considered [62].

It is widely accepted that the use of DSMLs may boost the modeling practice and enhance the flexibility, maintainability, and sustained use of a model compared with using a GPML [74,75]. Developing a DSML is very costly and time consuming. It is premature to declare the impact and possibility of continued use of all the included DSMLs in this review because of their recent appearances [65,66]. However, the justifications for embarking on such a time-consuming and costly process need to be recognized to gain inputs to build the most expressive and simplified modeling language for the domain. For example, TP-VML aimed at replacing all the extensions to GPMLs that are not built to represent all the needs in the domain [65].

On the basis of the anatomy of visual modeling languages [43], a visual vocabulary (graphical symbols), grammar (a set of compositional rules), semantics (definitions of the meaning of each symbol), and its visual (concrete) syntax are included. Most articles included in this review did not specifically aim at the core aspects of a modeling language by discussing its grammar, ontologies, and semantics fully. As this scoping review aims at mapping but not quality appraising the modeling languages, we did not seek out further information for those

developed or traced back to the original specifications for those adopted, adapted, or combined known visual modeling languages. There are frameworks with comprehensive evaluation criteria to measure the quality of a modeling language, although the criteria confuse the script with grammar [44]. A few papers included in this review also presented their criteria for selecting one modeling language over another. Future studies need to find a good framework and evaluate all modeling languages used in PP in general and DSMLs in particular. As clearly recommended by Gemino and Wand [73], a comparison of modeling languages should be based on their grammar (constructs and rules) rather than the scripts (specific models and end product of a modeling process).

The modeling and selection of the appropriate modeling language go in line with the data sources. Aspland et al [42] stated that there are 2 common ways of obtaining data: either data-driven approaches or through collaboration with those who interact with the pathway. They recommended coordinating both sources to advance on the advantages of each. The increasing presence of digital technology presents the opportunity to use a data-driven approach, which also gives the opportunity to evaluate progress in a dynamic manner [12,33,40]. More emphasis is being placed on the presence of an EHR to have access to reliable data at all phases to "identify the relationships between the context, the mechanisms, and the results obtained" [12]. Of note is the extensive use of other modeling languages that are more relevant to the modeling methods categorized as stochastic, data mining and machine learning, simulation, and optimization and heuristics by Aspland et al [42]. Our review encompasses the use of modeling languages for pathways created by multidisciplinary teams in a prospective manner.

With the increasing use of patient-centric pathway development [12], modeling languages that reflect the patient perspectives are of interest. The patient journey concept, or customer and user journey in a wider context, is getting more attention; although, the literature on the subject is incoherent [76]. One such language that covers the perspective of patients is the customer journey modeling language (CJML), which is making its way to the health domain [77] targeting novice modelers [78]. CJML is developed from the human end user's perspective, as opposed to the software-centric UML and the business-centric BPMN (refer to the diagrams of CJML from one of the authors' previous works for illustration purposes in Multimedia Appendix 3). This modeling language has been explored to support the design, management, and analysis of patient journeys. CJML is designed to capture both planned (prescribed) and actual patient journeys, allowing for the analysis of deviations and the inclusion of the end-users' experience [79]. CJML provides vocabulary, a metamodel, and a visual notation. It is easy to learn and does not require complex technical competence to use. CJML provides a graphical notation for the planned and actual journeys with specific constructs such as journey phases, experience, channel, and actor information, which is limited or lacking in other process-modeling languages [77-79].

### Limitations of This Scoping Review

We restricted our inclusion criteria based on several factors. Although this has merits to focus this review on modeling languages that emphasized description and communication of PPs, our narrow focus, including restrictions of the search strategies to the title and keywords only, might have excluded papers that have relevance regarding how the modeling languages were used to address domain issues while used in the modeling of data-driven pathways. Our effort to include DSMLs, which are not yet standardized, posed a challenge as to where to draw the line in the abundance of modeling languages [38,74] with varying levels of abstraction and granularity. We excluded simple flowchart and diagram presentations, the most common visual notations, because the scope of this review includes advanced languages. However, a few matured visual presentations with less structured description of the language might have been excluded. We did not include all the standards and accompanying modeling languages and methods that focus on standardizing interoperability between different IT systems, such as HL7 (Health Level Seven International), and textual languages that support international or interinstitutional coding to classify and standardize terminologies, such as SNOMED-CT [80]. We excluded the so-called task-based modeling languages commonly used to digitize clinical practice guidelines [81], because they lie outside of our inclusion criteria. This complexity is exacerbated in the overall confusion of the PP definitions, as also presented in the Result section of this review.

Identifying the operational definitions of the PP by the authors has been challenging. We filtered articles based on their emphasis on defining and explicitly stating where the focus of their work mainly targets. However, there is a fine line between closely similar concepts, such as workflow and clinical practice guidelines, which were understood as synonyms in a few original papers and previous reviews [12,37,42]. In contrast, even though terms that are used to represent similar concepts with PPs (those that were highlighting the organizational aspects of a care delivery) are presented in the title of the paper, the definition and the emphasis given to the concept in the body of the paper did not demonstrate the intention. These papers were excluded. We did not include patient journeys in the search strategy because it differs from the PP concept definition (in 1 review included synonymously [12]), which resulted in excluding modeling languages that are patient centric, such as CJML.

We have not conducted ontological analysis to check for the correspondence of ontologies and notations [28] or any other quality appraisal frameworks [73]. The use of such tools would have excluded a few papers that are more focused on scripts and are more method-intensive where limited information on the core modeling language and notations were presented. We have not stated a strict definition of a modeling language that implicates the need for a quality appraisal. Identifying the known and standardized modeling languages was simple. Applying strict operational definitions to those lesser-known DSMLs was complex and required careful analysis. Therefore, we considered the authors' description of the modeling language to accommodate in this scoping review. This resulted in the inclusion of papers focused on techniques and methods with

less information about the modeling languages that were used. Some raised important questions about a modeling language but used a more generic description in the part where they described the tool [68].

The variability in extractable data from each paper significantly influenced the focus of our discussion. Despite using the Wand and Weber framework [44], the lack of detailed information within each manuscript constrained our ability to engage in deeper discussions within this review. While the aim of this scoping review was not to conduct quality appraisal, the limitations in available data impeded our ability to thoroughly identify research gaps in a more nuanced manner.

### Gaps Identified for Future Research

Finding effective ways of describing and communicating a PP is essential to enhance its impact. It is as important as having a framework for its development and implementation. A considerable number of frameworks suggesting a standardized development of a PP to set common and shared practices do exist [12,27]. However, less focus has been given on how to simplify and optimize the description and communication of a model beyond the use of customary box-and-arrow type flowcharts and narrative text descriptions. Given the amount of locally made PPs, the effort directed toward the use of visual modeling language is very scarce. A report from PPs in oncology care research group [82] found that many of the cancer care pathways were presented in the form of flowcharts and texts. We also reviewed the national websites where the popular standardized cancer care pathways, for example, locally known as “pakkeforløp for kreft” in Norway, are presented in Scandinavian countries [83-85]. The pathways for several cancer cases were presented in a flowchart, a narrative text, and a table format. We have not explored how they are presented at each institution level, including the process of integration with EHR systems. As per the process steps of the cancer care pathways, the need for a standardized visual modeling language may not be argued provided that the intended goal is achieved by the current representations. This example and the overall limited adoption of visual modeling languages suggest that despite the inherent ambiguity and lack of precision and consistency associated with narrative and text descriptions [62], many institutions still rely on these representations over visual models. This reliance may stem from factors such as organizational culture, lack of awareness, and apprehension about the complexity of modeling languages. However, raising awareness and fostering a better understanding of the benefits and drawbacks of visual models, particularly in the context of specific PP projects, could help shift attitudes and reshape organizational culture regarding the choice of process representation.

Efforts to develop simplified visual modeling languages accessible to nonmodeler domain experts can help alleviate concerns about complexity and encourage broader adoption. While balancing ease of understanding with the expressiveness of modeling languages poses challenges [65], it is essential for all contributions in this field to prioritize this as a principle. In addition, as digital technology plays an increasingly prominent role in this domain, emphasis should be placed on visual languages equipped with execution engines and tools, facilitating integration with existing digital tools, and streamlining the modeling process.

In larger projects intended for widespread implementation, such as those aiming to serve as common documents for local implementations, incorporating visual models of business processes benefits the project while normalizing the use of such representations. For instance, the World Health Organization’s smart guideline documents have integrated visual representations of business processes [86]. However, further research is needed to thoroughly evaluate the advantages and disadvantages of visually modeling such projects using an accessible and expressive modeling language.

Using the decision model to select the most appropriate process-modeling language for a given modeling task is proposed for research modelers, which can be applied in the PP domain as well [74]. Guizani and Ghannouchi [75] argued that none of the languages that they reviewed fully supported the following 7 criteria: expressiveness, flexibility, formality, readability, support tools, usability, and ease of learning. Therefore, multicriteria decision support analysis is suggested as the most appropriate approach for comparison. We recommend that modelers and PP developers should exhaustively scrutinize the suitability of existing standards to model their PP before embarking on developing a DSML from scratch. However, reviews are needed to avail comprehensive knowledge of the gaps in the existing potential candidate languages and weigh the merits of developing new ones in relation to whether all requirements in a PP concept are covered or not.

Extension of existing languages, especially extension by addition, is a good approach to cover domain requirements but the standardization process of the extensions is a challenge [39,49]. By contrast, the latest OMG-recommended combinations of modeling languages have to be tested in different contexts as it is a promising approach to cover all the domain needs [62]. With the move toward a more patient-centric nature of pathways, more research is needed on how to reflect the patients’ views in the modeling product [12,25,87].

## Conclusions

Diverse visual modeling languages were used to model PPs fully or partially. The GPMLs were used directly without any modification to the grammar, extended following extension protocols, or combined with other languages to complement the inherent limitations of each language. We identified a few attempts of developing DSMLs in this review. A limited number of papers presented a DSML that is developed to meet the specific requirements of the PPs. Purely ontological modeling languages were also identified. We have shown the need to consider the rationale, context, and the ways in which the identified visual modeling languages were used. This provides additional useful information to stakeholders in the process of selecting a modeling language. Furthermore, one should use quality appraisal tools to check the conformity of a modeling language to their specific pathway project before deciding to use, extend, combine, or develop a visual modeling language.

## Acknowledgments

The authors would like to thank the Research Council of Norway for funding the Pathway project (grant 316342) and the members and contributors of the Pathway project.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 16 KB - ijmr_v13i1e55865_app1.docx ]

Multimedia Appendix 2
Context in which the modeling languages are applied.
[DOC File , 69 KB - ijmr_v13i1e55865_app2.doc ]

Multimedia Appendix 3
Example diagrams of the customer journey modeling language, adopted from one of the authors' previous work for illustration purposes; customer journey diagram (upper part): in this case, a patient consulting a general practitioner and being referred to a specialist. The swimlane diagram (lower part) reveals the network of actors involved in the patient's journey).
[PNG File , 184 KB - ijmr_v13i1e55865_app3.png ]

Multimedia Appendix 4
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 498 KB - ijmr_v13i1e55865_app4.pdf ]

## References

1. Seys D, Coeckelberghs E, Sermeus W, Van Zelm R, Panella M, Babu Payedimarri A, et al. Overview on the target population and methods used in care pathway projects: a narrative review. Int J Clin Pract 2021 Oct;75(10):e14565. [doi: 10.1111/ijcp.14565] [Medline: 34165865]
2. Adjemian R, Moradi Zirkohi A, Coombs R, Mickan S, Vaillancourt C. Are emergency department clinical pathway interventions adequately described, and are they delivered as intended? A systematic review. Int J Care Coord 2017 Sep 27;20(4):148-161. [doi: 10.1177/2053434517732507]
3. Khan A, Arthurs E, Gradin S, MacKinnon M, Sussman J, Kukreti V. Integrated care planning for cancer patients: a scoping review. Int J Integr Care 2017 Nov 13;17(6):5 [FREE Full text] [doi: 10.5334/ijic.2543] [Medline: 29588638]
4. Chawla A, Westrich K, Dai A, Mantels S, Dubois RW. US care pathways: continued focus on oncology and outstanding challenges. Am J Manag Care 2019 Jun;25(6):280-287 [FREE Full text] [Medline: 31211555]
5. Xu G, Yang Y, Du Y, Peng F, Hu P, Wang R, et al. Clinical pathway for early diagnosis of COVID-19: updates from experience to evidence-based practice. Clin Rev Allergy Immunol 2020 Aug;59(1):89-100 [FREE Full text] [doi: 10.1007/s12016-020-08792-8] [Medline: 32328954]
6. Trimarchi L, Caruso R, Magon G, Odone A, Arrigoni C. Clinical pathways and patient-related outcomes in hospital-based settings: a systematic review and meta-analysis of randomized controlled trials. Acta Biomed 2021 Feb 09;92(1):e2021093 [FREE Full text] [doi: 10.23750/abm.v92i1.10639] [Medline: 33682818]
7. Asmirajanti M, Syuhaimie Hamid AY, Hariyati TS. Clinical care pathway strenghens interprofessional collaboration and quality of health service: a literature review. Enferm Clin 2018 Feb;28:240-244. [doi: 10.1016/S1130-8621(18)30076-7]
8. Dunnack HJ, Montano AR. Interprofessional clinical pathway program effects on patient outcomes in the setting of neutropenic fever: an integrative review. Eur J Oncol Nurs 2021 Jun;52:101974. [doi: 10.1016/j.ejon.2021.101974] [Medline: 33991870]
9. Riza RC, Nurwahyuni A. The implementation and outcome of clinical pathway: a systematic review. In: Proceedings of the 5th International Conference on Public Health. 2019 Presented at: ICPH 2019; February 13-14, 2019; Solo, Indonesia URL: https://theicph.com/wp-content/uploads/2019/07/33.-Rizka-Chairani-Riza.pdf [doi: 10.26911/theicph.2019.05.33]
10. Rotter T, Kugler J, Koch R, Gothe H, Twork S, van Oostrum JM, et al. A systematic review and meta-analysis of the effects of clinical pathways on length of stay, hospital costs and patient outcomes. BMC Health Serv Res 2008 Dec 19;8(1):265 [FREE Full text] [doi: 10.1186/1472-6963-8-265] [Medline: 19094244]

11. van Hoeve JC, Vernooij RW, Fiander M, Nieboer P, Siesling S, Rotter T. Effects of oncological care pathways in primary and secondary care on patient, professional and health systems outcomes: a systematic review and meta-analysis. Syst Rev 2020 Oct 25;9(1):246 [FREE Full text] [doi: 10.1186/s13643-020-01498-0] [Medline: 33100227]
12. Gartner JB, Abasse KS, Bergeron F, Landa P, Lemaire C, Côté A. Definition and conceptualization of the patient-centered care pathway, a proposed integrative framework for consensus: a concept analysis and systematic review. BMC Health Serv Res 2022 Apr 26;22(1):558 [FREE Full text] [doi: 10.1186/s12913-022-07960-0] [Medline: 35473632]
13. Baker E, Woolley A, Xyrichis A, Norton C, Hopkins P, Lee G. How does the implementation of a patient pathway-based intervention in the acute care of blunt thoracic injury impact on patient outcomes? A systematic review of the literature. Injury 2020 Aug;51(8):1733-1743 [FREE Full text] [doi: 10.1016/j.injury.2020.06.002] [Medline: 32576379]
14. Seys D, Panella M, VanZelm R, Sermeus W, Aeyels D, Bruyneel L, et al. Care pathways are complex interventions in complex systems: New European Pathway Association framework. Int J Care Coord 2019 Mar 20;22(1):5-9. [doi: 10.1177/2053434519839195]
15. Latina R, Salomone K, D'Angelo D, Coclite D, Castellini G, Gianola S, et al. Towards a new system for the assessment of the quality in care pathways: an overview of systematic reviews. Int J Environ Res Public Health 2020 Nov 20;17(22):8634 [FREE Full text] [doi: 10.3390/ijerph17228634] [Medline: 33233824]
16. Chawla A, Westrich K, Matter S, Kaltenboeck A, Dubois R. Care pathways in US healthcare settings: current successes and limitations, and future challenges. Am J Manag Care 2016 Jan;22(1):53-62 [FREE Full text] [Medline: 26799125]
17. Adisandiya I, Savitri M. Barrier and enabler of clinical pathway implementation in hospital: a systematic review. In: Proceedings of the 5th International Conference on Public Health. 2019 Presented at: ICPH 2019; February 13-14, 2019; Solo, Indonesia URL: https://theicph.com/wp-content/uploads/2019/07/45.-Indiraya-Adisandiya1.pdf [doi: 10.26911/theicph.2019.04.45]
18. de Luc K, Kitchiner D. Developing Care Pathways, Volume 1. London, UK: Radcliffe Medical Press; 2001.
19. de Bleser L, Depreitere R, de Waele K, Vanhaecht K, Vlayen J, Sermeus W. Defining pathways. J Nurs Manag 2006 Oct;14(7):553-563. [doi: 10.1111/j.1365-2934.2006.00702.x] [Medline: 17004966]
20. Vanhaecht K. The impact of clinical pathways on the organisation of care processes. ResearchGate. 2007 Oct 2. URL: https://www.researchgate.net/publication/228996808_The_impact_of_clinical_pathways_on_the_organisation_of_care_processes [accessed 2023-11-01]
21. Kinsman L, Rotter T, James E, Snow P, Willis J. What is a clinical pathway? Development of a definition to inform the debate. BMC Med 2010 May 27;8:31 [FREE Full text] [doi: 10.1186/1741-7015-8-31] [Medline: 20507550]
22. Lawal AK, Rotter T, Kinsman L, Machotta A, Ronellenfitsch U, Scott SD, et al. What is a clinical pathway? Refinement of an operational definition to identify clinical pathway studies for a Cochrane systematic review. BMC Med 2016 Feb 23;14:35 [FREE Full text] [doi: 10.1186/s12916-016-0580-z] [Medline: 26904977]
23. Adjemian R, Zirkohi AM, Coombs R, Mickan S, Vaillancourt C. Validation of descriptive clinical pathway criteria in the systematic identification of publications in emergency medicine. Int J Care Coord 2017 May 08;20(1-2):45-56. [doi: 10.1177/2053434517707971]
24. Schrijvers G, van Hoorn A, Huiskes N. The care pathway: concepts and theories: an introduction. Int J Integr Care 2012 Jan;12(Spec Ed Integrated Care Pathways):e192 [FREE Full text] [doi: 10.5334/ijic.812] [Medline: 23593066]
25. Davies EL, Bulto LN, Walsh A, Pollock D, Langton VM, Laing RE, et al. Reporting and conducting patient journey mapping research in healthcare: a scoping review. J Adv Nurs 2023 Jan 03;79(1):83-100 [FREE Full text] [doi: 10.1111/jan.15479] [Medline: 36330555]
26. Seguin ML, Rangnekar A, Renedo A, Palafox B, McKee M, Balabanova D. Systematic review of frameworks used to conceptualise health pathways of individuals diagnosed with cardiovascular diseases. BMJ Glob Health 2020 Sep;5(9):e002464 [FREE Full text] [doi: 10.1136/bmjgh-2020-002464] [Medline: 32928800]
27. Flores EJ, Mull NK, Lavenberg JG, Mitchell MD, Leas BF, Williams A, et al. Using a 10-step framework to support the implementation of an evidence-based clinical pathways programme. BMJ Qual Saf 2019 Jun;28(6):476-485. [doi: 10.1136/bmjqs-2018-008454] [Medline: 30463885]
28. Recker J, Rosemann M, Krogstie J. Ontology- versus pattern-based evaluation of process modeling languages: a comparison. Commun Assoc Inf Syst 2007;20(48) [FREE Full text] [doi: 10.17705/1CAIS.02048]
29. Lu R, Sadiq S. A survey of comparative business process modeling approaches. In: Proceedings of the 10th International Conference on Business Information Systems. 2007 Presented at: BIS 2007; April 25-27, 2007; Poznan, Poland. [doi: 10.1007/978-3-540-72035-5_7]
30. Dumas M, La Rosa M, Mendling J, Reijers HA. Fundamentals of Business Process Management. Berlin, Heidelberg: Springer; 2018.
31. Recker J, Rosemann M, Indulska M, Green P. Business process modeling- a comparative analysis. J Assoc Inf Syst 2009 Apr;10(04):333-363. [doi: 10.17705/1jais.00193]
32. Askari M, Tam JL, Klundert J. The effectiveness of clinical pathway software in inpatient settings: a systematic review. Int J Med Inform 2021 Mar;147:104374 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104374] [Medline: 33422761]

33. Neame MT, Chacko J, Surace AE, Sinha IP, Hawcutt DB. A systematic review of the effects of implementing clinical pathways supported by health information technologies. J Am Med Inform Assoc 2019 Apr 01;26(4):356-363 [FREE Full text] [doi: 10.1093/jamia/ocy176] [Medline: 30794311]
34. Scheplitz T. Pathway-supporting health information systems: a review. Stud Health Technol Inform 2021 May 07;279:87-94. [doi: 10.3233/SHTI210093] [Medline: 33965923]
35. Elbattah M, Molloy O. Towards improving modeling and simulation of clinical pathways: lessons learned and future insights. In: Proceedings of the 5th International Conference on Simulation and Modeling Methodologies, Technologies and Applications. 2015 Presented at: SIMULTECH 2015; July 21-23, 2015; Colmar, France. [doi: 10.5220/0005568405080514]
36. Yang W, Su Q. Process mining for clinical pathway: literature review and future directions. In: Proceedings of the 11th International Conference on Service Systems and Service Management. 2014 Presented at: ICSSSM 2014; June 25-27, 2014; Beijing, China. [doi: 10.1109/icsssm.2014.6943412]
37. Mincarone P, Leo CG, Trujillo-Martín MD, Manson J, Guarino R, Ponzini G, et al. Standardized languages and notations for graphical modelling of patient care processes: a systematic review. Int J Qual Health Care 2018 Apr 01;30(3):169-177. [doi: 10.1093/intqhc/mzx197] [Medline: 29346638]
38. Mili H, Tremblay G, Jaoude GB, Lefebvre É, Elabed L, Boussaidi GE. Business process modeling languages: sorting through the alphabet soup. ACM Comput Surv 2010 Dec 03;43(1):1-56. [doi: 10.1145/1824795.1824799]
39. Pufahl L, Zerbato F, Weber B, Weber I. BPMN in healthcare: challenges and best practices. Inf Syst 2022 Jul;107:102013. [doi: 10.1016/j.is.2022.102013]
40. Combi C, Oliboni B, Gabrieli A. Conceptual modeling of clinical pathways: making data and processes connected. In: Proceedings of the 15th Conference on Artificial Intelligence in Medicine. 2015 Presented at: AIME 2015; June 17-20, 2015; Pavia, Italy. [doi: 10.1007/978-3-319-19551-3_7]
41. Pahl C, Supriyanto E, Detschew V. Evaluation of UML modeling tools for clinical pathways used in health information systems. In: Proceeding of the 10th Iberian Conference on Information Systems and Technologies. 2015 Presented at: CISTI 2015; June 17-20, 2015; Aveiro, Portugal. [doi: 10.1109/cisti.2015.7170540]
42. Aspland E, Gartner D, Harper P. Clinical pathway modelling: a literature review. Health Syst (Basingstoke) 2019 Sep 11;10(1):1-23 [FREE Full text] [doi: 10.1080/20476965.2019.1652547] [Medline: 33758656]
43. Moody D. The “physics” of notations: toward a scientific basis for constructing visual notations in software engineering. IIEEE Trans Softw Eng 2009 Nov;35(6):756-779. [doi: 10.1109/tse.2009.67]
44. Wand Y, Weber R. Research commentary: information systems and conceptual modeling—a research agenda. Inf Syst Res 2002 Dec;13(4):363-376. [doi: 10.1287/isre.13.4.363.69]
45. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
46. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
47. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
48. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
49. Zarour K, Benmerzoug D, Guermouche N, Drira K. A systematic literature review on BPMN extensions. BPMJ 2019 Nov 18;26(6):1473-1503. [doi: 10.1108/bpmj-01-2019-0040]
50. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2024-09-29]
51. Scheuerlein H, Rauchfuss F, Dittmar Y, Molle R, Lehmann T, Pienkos N, et al. New methods for clinical pathways-Business Process Modeling Notation (BPMN) and Tangible Business Process Modeling (t.BPM). Langenbecks Arch Surg 2012 Jun;397(5):755-761. [doi: 10.1007/s00423-012-0914-z] [Medline: 22362053]
52. Barbagallo S, Corradi L, de Ville de Goyet J, Iannucci M, Porro I, Rosso N, et al. Optimization and planning of operating theatre activities: an original definition of pathways and process modeling. BMC Med Inform Decis Mak 2015 May 17;15(1):38 [FREE Full text] [doi: 10.1186/s12911-015-0161-7] [Medline: 25982033]
53. Ferrante S, Bonacina S, Pinciroli F. Modeling stroke rehabilitation processes using the Unified Modeling Language (UML). Comput Biol Med 2013 Oct;43(10):1390-1401. [doi: 10.1016/j.compbiomed.2013.07.012] [Medline: 24034730]
54. Ferrante S, Bonacina S, Pozzi G, Pinciroli F, Marceglia S. A design methodology for medical processes. Appl Clin Inform 2017 Dec 16;07(01):191-210. [doi: 10.4338/aci-2015-08-ra-0111]
55. Mauro C, Happle T, Sunyaev A, Leimeister JM, Krcmar H. From medical processes to workflows: modeling of clinical pathways with the unified modeling language. In: Proceedings of the Third International Conference on Health Informatics. 2010 Presented at: HEALTHINF 2010; January 20-23, 2010; Valencia, Spain.
56. Zerbato F, Oliboni B, Carlo C, Campos M, Juarez JM. BPMN-based representation and comparison of clinical pathways for catheter-related bloodstream infections. In: Proceedings of the International Conference on Healthcare Informatics. 2015 Presented at: ICHI 2015; October 21-23, 2015; Dallas, TX. [doi: 10.1109/ichi.2015.49]

57. Braun R, Schlieter H, Burwitz M, Esswein W. BPMN4CP revised -- extending BPMN for multi-perspective modeling of clinical pathways. In: Proceedings of the 49th Hawaii International Conference on System Sciences. 2016 Presented at: HICSS 2016; January 05-08, 2016; Koloa, HI. [doi: 10.1109/hicss.2016.407]
58. Richter P, Schlieter H. Process-based quality management in care: adding a quality perspective to pathway modelling. In: Proceedings of the OTM Confederated International Conferences "On the Move to Meaningful Internet Systems". 2019 Presented at: OTM 2019; October 21-25, 2019; Rhodes, Greece. [doi: 10.1007/978-3-030-33246-4_25]
59. Tehrani J, Liu K, Michell V. Semiotics-oriented method for generation of clinical pathways. In: Proceedings of 2nd International Conference on Logistics, Informatics and Service Science. 2012 Presented at: LISS 2012; July 12-15, 2012; Beijing, China. [doi: 10.1007/978-3-642-32054-5_69]
60. Combi C, Oliboni B, Zardini A, Zerbato F. A methodological framework for the integrated design of decision-intensive care pathways-an application to the management of COPD patients. J Healthc Inform Res 2017 Dec 17;1(2):157-217 [FREE Full text] [doi: 10.1007/s41666-017-0007-4] [Medline: 35415395]
61. Sooter LJ, Hasley S, Lario R, Rubin KS, Hasić F. Modeling a clinical pathway for contraception. Appl Clin Inform 2019 Oct;10(5):935-943 [FREE Full text] [doi: 10.1055/s-0039-3400749] [Medline: 31860113]
62. Field guide to shareable clinical pathways. Object Management Group Healthcare Domain Taskforce. 2020 Jan 01. URL: https://www.trisotech.com/wp-content/uploads/OMG-Field-Guide-to-sharable-Clinical-Pathways.pdf [accessed 2023-10-12]
63. Bowles J, Caminati MB, Cha S. An integrated framework for verifying multiple care pathways. In: Proceedings of the International Symposium on Theoretical Aspects of Software Engineering. 2017 Presented at: TASE 2017; September 13-15, 2017; Sophia Antipolis, France. [doi: 10.1109/tase.2017.8285628]
64. Ardito C, Caivano D, Colizzi L, Dimauro G, Verardi L. Design and execution of integrated clinical pathway: a simplified meta-model and associated methodology. Information 2020 Jul 13;11(7):362. [doi: 10.3390/info11070362]
65. Iglesias N, Juarez JM, Campos M. Business process model and notation and openEHR task planning for clinical pathway standards in infections: critical analysis. J Med Internet Res 2022 Sep 15;24(9):e29927 [FREE Full text] [doi: 10.2196/29927] [Medline: 36107480]
66. Trajano IA, Ferreira Filho JB, de Carvalho Sousa FR, Litchfield I, Weber P. MedPath: a process-based modeling language for designing care pathways. Int J Med Inform 2021 Feb;146:104328. [doi: 10.1016/j.ijmedinf.2020.104328] [Medline: 33281069]
67. Burwitz M, Schlieter H, Esswein W. Modeling clinical pathways - design and application of a domain-specific modeling language. In: Proceedings of the 11th International Conference on Wirtschaftsinformatik. 2013 Presented at: WI2013; February 27-March 1, 2013; Leipzig, Germany URL: https://aisel.aisnet.org/wi2013/83
68. Shitkova M, Taratukhin V, Becker J. Towards a methodology and a tool for modeling clinical pathways. Procedia Comput Sci 2015;63:205-212. [doi: 10.1016/j.procs.2015.08.335]
69. Li W, Liu K, Li S, Yang H. Normative modeling for personalized clinical pathway using organizational semiotics methods. In: Proceedings of the International Symposium on Computer Science and Computational Technology. 2008 Presented at: ISCSCT 2008; December 20-22, 2008; Shanghai, China. [doi: 10.1109/iscsct.2008.310]
70. Ye Y, Jiang Z, Yang D, Du G. A semantics-based clinical pathway workflow and variance management framework. In: Proceedings of the IEEE International Conference on Service Operations and Logistics, and Informatics. 2008 Presented at: SOLI 2008; October 12-15, 2008; Beijing, China. [doi: 10.1109/soli.2008.4686499]
71. Nishimura S, Nishijima G, Kitamura Y, Sasajima M, Takeda T, Matsumura Y, et al. CHARMing clinical pathways modeling of clinical pathways based on the goal-oriented ontological framework CHARM. In: Proceedings of the International Conference on Health Informatics. 2014 Presented at: HEALTHINF-2014; March 3-6, 2014; Loire Valley, France. [doi: 10.5220/0004747002110221]
72. Abidi SR, Abidi SS. An ontological modeling approach to align institution-specific clinical pathways: towards inter-institution care standardization. In: Proceedings of the 25th IEEE International Symposium on Computer-Based Medical Systems. 2012 Presented at: CBMS 2012; June 20-22, 2012; Rome, Italy. [doi: 10.1109/cbms.2012.6266392]
73. Gemino A, Wand Y. A framework for empirical evaluation of conceptual modeling techniques. Requir Eng 2004 Oct 15;9:248-260. [doi: 10.1007/s00766-004-0204-6]
74. Farshidi S, Kwantes IB, Jansen S. Business process modeling language selection for research modelers. Softw Syst Model 2023 May 29;23:137-162. [doi: 10.1007/s10270-023-01110-8]
75. Guizani K, Ghannouchi SA. An approach for selecting a business process modeling language that best meets the requirements of a modeler. Procedia Comput Sci 2021;181:843-851. [doi: 10.1016/j.procs.2021.01.238]
76. Tueanrat Y, Papagiannidis S, Alamanos E. Going on a journey: a review of the customer journey literature. J Bus Res 2021 Mar;125:336-353. [doi: 10.1016/j.jbusres.2020.12.028]
77. Halvorsrud R, Lund Lillegaard A, Røhne M, Jensen AM. Managing complex patient journeys in healthcare. In: Pfannstiel MA, Rasche C, editors. Service Design and Service Thinking in Healthcare and Hospital Management. Cham, Switzerland: Springer; 2019.
78. Halvorsrud R, Haugstveit IM, Pultier A. Evaluation of a modelling language for customer journeys. In: Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing. 2016 Presented at: VLHCC 2016; September 4-8, 2016; Cambridge, UK. [doi: 10.1109/vlhcc.2016.7739662]

79. Halvorsrud R, Sanchez OR, Boletsis C, Skjuve M. Involving users in the development of a modeling language for customer journeys. Softw Syst Model 2023 Jan 19;22:1589-1618. [doi: 10.1007/s10270-023-01081-w]
80. Alahmar A, Crupi ME, Benlamri R. Ontological framework for standardizing and digitizing clinical pathways in healthcare information systems. Comput Methods Programs Biomed 2020 Nov;196:105559. [doi: 10.1016/j.cmpb.2020.105559] [Medline: 32531654]
81. Peleg M. Computer-interpretable clinical guidelines: a methodological review. J Biomed Inform 2013 Aug;46(4):744-763 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.009] [Medline: 23806274]
82. Richter P, Schlieter H. Patient pathways in oncology care results of a systematic literature review. Innovative Partnership for Action Against Cancer. 2021 Mar 8. URL: https://www.ipaac.eu/res/file/outputs/wp10/patient-pathways-oncology-care.pdf [accessed 2024-09-30]
83. Nasjonale anbefalinger, råd, pakkeforløp og pasientforløp. Helsedirektoratet. URL: https://www.helsedirektoratet.no/produkter [accessed 2023-12-12]
84. Standardiserade vårdförlopp i cancervården (SVF). Regionala Cancercentrum. URL: https://cancercentrum.se/samverkan/vara-uppdrag/kunskapsstyrning/vardforlopp/ [accessed 2023-12-19]
85. Probst HB, Hussain ZB, Andersen O. Cancer patient pathways in Denmark as a joint effort between bureaucrats, health professionals and politicians--a national Danish project. Health Policy 2012 Apr;105(1):65-70. [doi: 10.1016/j.healthpol.2011.11.001] [Medline: 22136810]
86. Mehl G, Tunçalp Ö, Ratanaprayul N, Tamrat T, Barreix M, Lowrance D, et al. WHO SMART guidelines: optimising country-level use of guideline recommendations in the digital age. Lancet Digit Health 2021 Apr;3(4):e213-e216 [FREE Full text] [doi: 10.1016/S2589-7500(21)00038-8] [Medline: 33610488]
87. Wind A, van der Linden C, Hartman E, Siesling S, van Harten W. Patient involvement in clinical pathway development, implementation and evaluation - a scoping review of international literature. Patient Educ Couns 2022 Jun;105(6):1441-1448 [FREE Full text] [doi: 10.1016/j.pec.2021.10.007] [Medline: 34666931]


## Abbreviations

**BPM:** business process management
**BPMN:** business process model and notation
**BPMN4CP:** business process model and notation for care pathways
**CJML:** customer journey modeling language
**CMMN:** case management model and notation
**DMN:** decision model and notation
**DSML:** domain-specific modeling language
**EHR:** electronic health record
**GPML:** general purpose–modeling language
**HL7:** Health Level Seven International
**OMG:** Object Management Group
**PP:** patient pathway
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
**SNOMED-CT:** Systematized Nomenclature of Medicine–Clinical Terms
**TP:** task planning
**TP-VML:** task planning visual modeling language
**UML:** unified modeling language


*Edited by T de Azevedo Cardoso; submitted 27.12.23; peer-reviewed by A Hassan, R Bidkar, GK Gupta; comments to author 24.04.24; revised version received 15.05.24; accepted 29.08.24; published 15.11.24.*

*Please cite as:*
*Bogale B, Vesinurm M, Lillrank P, Celius EG, Halvorsrud R*
*Visual Modeling Languages in Patient Pathways: Scoping Review*
*Interact J Med Res 2024;13:e55865*
*URL: https://www.i-jmr.org/2024/1/e55865*
*doi:10.2196/55865*
*PMID:*

Original Paper

# Evaluation of an mHealth App on Self-Management of Osteoporosis: Prospective Survey Study

Magnus Grønlund Bendtsen[1], MSc; Bodil Marie Thuesen Schönwandt[1], MSc; Mette Rubæk[1,2], MSc; Mette Friberg Hitz[1,3], PhD

[1]Research Unit, Medical Department, Zealand University Hospital, Koege, Denmark

[2]University of Southern Denmark, Odense, Denmark

[3]Institute 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), 4.5 (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.

***(Interact J Med Res 2024;13:e53995)*** 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 Zealand 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 ($T \leq -2.5$ 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 1.** Participant characteristics (N=95).

| Characteristics | Participants |
|---|---|
| Age (years), mean (SD) | 66.3 (8.4) |
| **Sex, n (%)** | |
| Male | 5 (5) |
| Female | 90 (95) |
| **Shorter education, n (%)** | 42 (44) |
| Primary school | 8 (8) |
| Vocational training, high school | 21 (22) |
| Higher education—short | 12 (14) |
| **Longer education, n (%)** | 53 (56) |
| Higher education—intermediate | 41 (43) |
| Higher education—long | 12 (13) |
| **Currently working, n (%)** | 33 (35) |
| Self-employed (professional) | 3 (3) |
| Civil servant | 23 (24) |
| Vocational | 3 (3) |
| Unskilled or semiskilled worker | 4 (4) |
| **Not currently working, n (%)** | 62 (65) |
| Unemployed | 3 (3) |
| Early retirement | 17 (18) |
| Retirement pension | 41 (43) |
| Leave of absence | 1 (1) |

### *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.

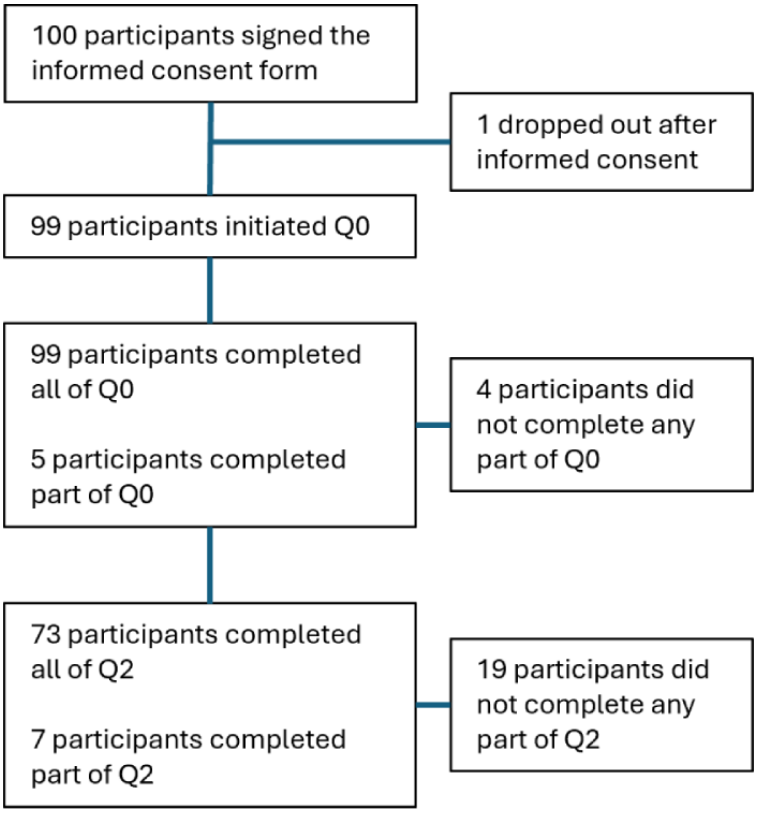



**Table 2.** *T* scores and fractures (N=95).

| Characteristics | Participants |
|---|---|
| Time since diagnosis(years), mean (range) | 6 (0-34) |
| ***T* score lumbar spine, n (%)** | |
| –2.5 to –3.0 | 24 (26) |
| –3.0 to –4.0 | 14 (15) |
| ≥–4.0 | 7 (7) |
| Do not know my *T* score | 49 (52) |
| ***T* score hip, n (%)** | |
| –2.5 to –3.0 | 26 (28) |
| –3.0 to –4.0 | 3 (3) |
| ≥–4.0 | 1 (1) |
| Do not know my *T* score | 64 (68) |
| **Previous fractures, n (%)** | 46 (49) |
| Wrist | 21 (22) |
| Upper arm | 3 (3) |
| Vertebrae | 15 (16) |
| Hip | 6 (6) |
| Other[a] | 19 (20) |

[a]Other fractures reported: fracture of foot, ankle, heel, toe, finger, elbow, and tibia.

**Table 3.** Medication (N=93).

| Characteristics | Participants |
|---|---|
| **Current treatment plan, n (%)** | 64 (67) |
| Oral BP[a] | 16 (25) |
| IV[b] BP | 26 (41) |
| Denosumab | 11 (17) |
| PTH[c] analog | 1 (2) |
| Romosozumab | 2 (3) |
| Unknown to me | 6 (9) |
| Other | 2 (3) |

[a]BP: bisphosphonate.

[b]IV: intravenous.

[c]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).

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

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

| Questionnaire | Value, mean (SD) |
|---|---|
| SUS | 74.6 (15.3) |
| **SUTAQ (scores from 1 to 6)** | |
| Domain 1 | 4.1 (1.2) |
| Domain 2 | 3.4 (1.1) |
| Domain 3 | 4.5 (1.1) |

**Table 6.** Service User Technology Acceptability Questionnaire individual statements (n=73).

| Statement | Agreement with statement, n (%) | | | | | |
|---|---|---|---|---|---|---|
| | 1[a] | 2[a] | 3[a] | 4[a] | 5[a] | 6[a] |
| **Domain 1** | | | | | | |
| The kit has allowed me to be less concerned about my health status. | 12 (16) | 3 (4) | 19 (26) | 13 (18) | 16 (22) | 10 (14) |
| The kit has made me more actively involved in my health. | 5 (7) | 2 (3) | 15 (21) | 21 (29) | 16 (22) | 14 (19) |
| The kit allows the people looking after me, to better monitor me and my condition. | 15 (21) | 14 (19) | 13 (18) | 11 (15) | 13 (18) | 7 (10) |
| The kit can be or should be recommended to people in a similar condition to mine. | 2 (3) | 1 (1) | 6 (8) | 7 (10) | 24 (33) | 33 (45) |
| The kit can certainly be a good addition to my regular health or social care. | 4 (5) | 1 (1) | 7 (10) | 17 (23) | 20 (27) | 24 (33) |
| **Domain 2** | | | | | | |
| 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. | 12 (16) | 6 (8) | 14 (19) | 18 (25) | 15 (21) | 8 (11) |
| The kit I received has increased my access to care (health or social care professionals) | 12 (16) | 7 (10) | 19 (26) | 22 (30) | 10 (14) | 3 (4) |
| The kit I received has helped me to improve my health | 6 (8) | 7 (10) | 13 (18) | 25 (34) | 16 (22) | 6 (8) |
| The kit has made it easier to get in touch with health and social care professionals. | 15 (21) | 8 (11) | 20 (27) | 16 (22) | 12 (16) | 2 (3) |
| **Domain 6** | | | | | | |
| The kit has been explained to me sufficiently. | 5 (7) | 10 (14) | 12 (16) | 12 (16) | 15 (21) | 19 (26) |
| The kit can be trusted to work appropriately. | 1 (1) | 1 (1) | 5 (7) | 11 (15) | 32 (44) | 23 (32) |
| I am satisfied with the kit I received. | 2 (3) | 3 (4) | 12 (16) | 7 (10) | 28 (38) | 21 (29) |

[a]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.

| Model | *R* | $R^2$ | Adjusted $R^2$ | SE of the estimate |
|---|---|---|---|---|
| 1 | 0.216[a] | 0.047 | 0.036 | 0.5046 |

[a]Predictors: (Constant). Occupation.

**Table 8.** ANOVA[a] for occupation's effect on eHealth Literacy Questionnaire domain 4.

| Model 1 | Sum of squares | *df* | Mean square | *F* (*df*) | *P* value |
|---|---|---|---|---|---|
| Regression | 1.099 | 1 | 1.099 | 4.315 (1,88) | .04[b] |
| Residual | 22.405 | 88 | 0.255 | __[c] | — |
| Total | 23.504 | 89 | — | — | — |

[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 Richtering 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 74.6 (15.3). According to Bangor et al [17], a score of approximately 70 indicates a user-friendliness rating of "Good." Conversely, scores around 35 and 50 correspond to ratings of "Poor" and "OK," respectively. On the other end of the spectrum, scores around 85 and 90 are associated with ratings of "Excellent" and "Best imaginable," respectively [17]. Based on the SUS score, the usability of the "My Bones" app is considered acceptable.

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 Hvidt (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

1. Kanis JA, Norton N, Harvey NC, Jacobson T, Johansson H, Lorentzon M, et al. SCOPE 2021: a new scorecard for osteoporosis in Europe. Arch Osteoporos 2021;16(1):82 [FREE Full text] [doi: 10.1007/s11657-020-00871-9] [Medline: 34080059]

2. Vestergaard P, Rejnmark L, Mosekilde L. Osteoporosis is markedly underdiagnosed: a nationwide study from Denmark. Osteoporos Int 2005;16(2):134-141 [FREE Full text] [doi: 10.1007/s00198-004-1680-8] [Medline: 15197546]
3. des Bordes J, Prasad S, Pratt G, Suarez-Almazor ME, Lopez-Olivo MA. Knowledge, beliefs, and concerns about bone health from a systematic review and metasynthesis of qualitative studies. PLoS One 2020;15(1):e0227765 [FREE Full text] [doi: 10.1371/journal.pone.0227765] [Medline: 31940409]
4. Hansen C, Konradsen H, Abrahamsen B, Pedersen BD. Women's experiences of their osteoporosis diagnosis at the time of diagnosis and 6 months later: a phenomenological hermeneutic study. Int J Qual Stud Health Well-being 2014;9:22438 [FREE Full text] [doi: 10.3402/qhw.v9.22438] [Medline: 24559545]
5. Hansen CA, Abrahamsen B, Konradsen H, Pedersen BD. Women's lived experiences of learning to live with osteoporosis: a longitudinal qualitative study. BMC Womens Health 2017;17(1):17 [FREE Full text] [doi: 10.1186/s12905-017-0377-z] [Medline: 28279157]
6. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016;39(6):981-994 [FREE Full text] [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
7. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
8. Slomian J, Appelboom G, Ethgen O, Reginster JY, Bruyère O. Can new information and communication technologies help in the management of osteoporosis? Womens Health (Lond) 2014;10(3):229-232 [FREE Full text] [doi: 10.2217/whe.14.15] [Medline: 24956289]
9. Alhussein G, Hadjileontiadis L. Digital health technologies for long-term self-management of osteoporosis: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(4):e32557 [FREE Full text] [doi: 10.2196/32557] [Medline: 35451968]
10. Jakobsen PR, Hermann AP, Søndergaard J, Wiil UK, Clemensen J. Development of an mHealth application for women newly diagnosed with osteoporosis without preceding fractures: a participatory design approach. Int J Environ Res Public Health 2018;15(2):330 [FREE Full text] [doi: 10.3390/ijerph15020330] [Medline: 29438343]
11. Jakobsen PR, Hermann AP, Søndergaard J, Wiil UK, Dixon RF, Clemensen J. Left in limbo—experiences and needs among postmenopausal women newly diagnosed with osteoporosis without preceding osteoporotic fractures: a qualitative study. Post Reprod Health 2018;24(1):26-33. [doi: 10.1177/2053369118755189] [Medline: 29392982]
12. Jakobsen PR, Hermann AP, Søendergaard J, Wiil UK, Jensen CM, Clemensen J. The gap between women's needs when diagnosed with asymptomatic osteoporosis and what is provided by the healthcare system: a qualitative study. Chronic Illn 2021;17(1):3-16 [FREE Full text] [doi: 10.1177/1742395318815958] [Medline: 30525980]
13. Jakobsen PR, Hermann AP, Søndergaard J, Wiil UK, Clemensen J. Help at hand: women's experiences of using a mobile health application upon diagnosis of asymptomatic osteoporosis. SAGE Open Med 2018;6:2050312118807617 [FREE Full text] [doi: 10.1177/2050312118807617] [Medline: 30397471]
14. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
15. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395 [FREE Full text] [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
16. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London: Taylor & Francis; 1995:189-194.
17. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123.
18. Hvidt JCS, Christensen LF, Sibbersen C, Helweg-Jørgensen S, Hansen JP, Lichtenstein MB. Translation and validation of the system usability scale in a Danish mental health setting using digital technologies in treatment interventions. Int J Hum Comput Interact 2019;36(8):709-716. [doi: 10.1080/10447318.2019.1680922]
19. Hirani SP, Rixon L, Beynon M, Cartwright M, Cleanthous S, Selva A, et al. Quantifying beliefs regarding telehealth: development of the whole systems demonstrator service user technology acceptability questionnaire. J Telemed Telecare 2017;23(4):460-469. [doi: 10.1177/1357633X16649531] [Medline: 27224997]
20. Holt KA, Karnoe A, Overgaard D, Nielsen SE, Kayser L, Røder ME, et al. Differences in the level of electronic health literacy between users and nonusers of digital health services: an exploratory survey of a group of medical outpatients. Interact J Med Res 2019;8(2):e8423 [FREE Full text] [doi: 10.2196/ijmr.8423] [Medline: 30950809]
21. García-García D, Bazán MJA, Pérez-Rivas FJ. Factors influencing eHealth literacy among Spanish primary healthcare users: cross-sectional study. Int J Environ Res Public Health 2022;19(23):15497 [FREE Full text] [doi: 10.3390/ijerph192315497] [Medline: 36497572]

22. Cheng C, Elsworth G, Osborne RH. Validity evidence based on relations to other variables of the eHealth Literacy Questionnaire (eHLQ): Bayesian approach to test for known-groups validity. J Med Internet Res 2021;23(10):e30243 [FREE Full text] [doi: 10.2196/30243] [Medline: 34647897]
23. Tassy A, Törnfelt C. It-anvendelse i befolkningen 2022. Danmarks Statistik. 2023. URL: https://www.dst.dk/da/Statistik/nyheder-analyser-publ/Publikationer/VisPub?cid=44692 [accessed 2023-08-25]
24. Chang SJ, Jang SJ, Lee H, Kim H. Building on evidence to improve eHealth literacy in older adults: a systematic review. Comput Inform Nurs 2021;39(5):241-247. [doi: 10.1097/CIN.0000000000000674] [Medline: 32925185]
25. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020;39(2):141-150 [FREE Full text] [doi: 10.1177/0733464818807468] [Medline: 30353776]
26. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
27. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
28. Sandvei M, Kynde I, Lee A. Fronløber projektet: evaluering af tilfredshed hos borgere og organisatoriske erfaringer. Syddansk Universitet. 2013. URL: https://www.sdu.dk/~/media/Files/Om_SDU/Centre/CAST/PDF_filer/Frontloeber.pdf [accessed 2023-08-25]
29. Lee A, Sandvei M, Hosbond T, Petersen J, Christiansen KR. Klinisk Integreret Hjemmemonitorering (KIH) Slutrapportering til Fonden for Velfaerdsteknologi. Digitaliseringsstyrelsen. 2015. URL: https://findresearcher.sdu.dk/ws/files/115999662/KIH_Slutrapport.pdf [accessed 2024-02-24] [doi: 10.13140/RG.2.1.5081.1363]


## Abbreviations

**DXA:** dual-energy x-ray absorptiometry
**eHLF:** 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


*Edited by M Yousufuddin; submitted 26.10.23; peer-reviewed by P Nohl-Deryk, C Jacob; comments to author 24.01.24; revised version received 05.02.24; accepted 10.02.24; published 01.04.24.*

*Please cite as:*
*Bendtsen MG, Schönwandt BMT, Rubæk M, Hitz MF*
*Evaluation of an mHealth App on Self-Management of Osteoporosis: Prospective Survey Study*
*Interact J Med Res 2024;13:e53995*
*URL: https://www.i-jmr.org/2024/1/e53995*
*doi:10.2196/53995*
*PMID:38557362*

Original Paper

# Intramural Health Care Through Video Consultations and the Need for Referrals and Hospital Admissions: Retrospective Quantitative Subanalysis of an Evaluation Study

Katharina Schmalstieg-Bahr[1,2], MD; Miriam Giovanna Colombo[3], MPH; Roland Koch[3], MD; Joachim Szecsenyi[4], MD; Friedrich Völker[2]; Eva Elisabeth Blozik[2,5], MD; Martin Scherer[1,2], MD

[1]Department of General Practice and Primary Care, University Medical Center Eppendorf, Hamburg, Germany

[2]A+ Videoclinic GmbH, Gräfelfing, Germany

[3]Institute for General Practice and Interprofessional Care, University Hospital Tübingen, Tübingen, Germany

[4]Department of General Practice and Health Services Research, University Hospital Heidelberg, Heidelberg, Germany

[5]Institute 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, 20246
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.

*(Interact J Med Res 2024;13:e44906)* doi:10.2196/44906

**KEYWORDS**

intramural health care; prison; telemedicine; primary care; family medicine; referral; hospital admission; admission rate; intramural; penal; video consult; e-consult; remote care; virtual care; health care delivery; service delivery; health care system

## Introduction

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.

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

### 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.

| Prison | Characteristics | Number of prisoners (n=2499), n (%) | Number of physicians (n=12), n (%) | Number of nurses (n=69), n (%) |
|---|---|---|---|---|
| P1 | Male adults | 350 (14) | 2 (17) | 9 (13) |
| P2 | Female adults or adolescents | 350 (14) | 1 (8) | 6 (9) |
| P3 | Male adolescents | 395 (16) | 4 (33) | 9 (13) |
| P4 | Male adults | 52 (2) | 0 (0) | 3 (4) |
| P5 | Male adults | 772 (31) | 4 (33) | 22 (32) |
| P6 | Male adults | 580 (23) | 1 (8) | 20 (29) |

### 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).

Due to the small number of necessary referrals, it is hard to rank specialties based on the frequency of necessary referrals. The data showed mixed results with psychiatry as the most common specialty referred to (3/17, 18%), followed by urology (2/17, 12%) and proctology (2/17, 12%). In 2 cases, patients were referred to an external psychiatrist; and 1 patient had a follow-up appointment with a prison psychiatrist. Referral diagnoses also varied between the cases (Table 3). A total of 5 out of the 12 VC physicians (4 family physicians and 1 psychiatrist) conducted the encounters resulting in a referral, however, the number of overall encounters per physician varied greatly ranging from 2 to 143. An interpreter was not needed in any of the 17 consultations that resulted in a referral.

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

| Type of consultation | Medical specialty (physician number) | Diagnosis | Medical specialty referred to |
|---|---|---|---|
| Unscheduled | FM[a] (D7) | Drug abuse, opiate withdrawal | Psychiatry (coming to site) |
| Scheduled | PSY[b] (D8) | Adjustment disorder | Child and adolescent psychiatry |
| Unscheduled | FM (D9) | Acneiform dermatitis | Dermatology (video consultation) |
| Scheduled | FM (D5) | Nausea, suspected adverse reaction, suspected melena | Rheumatology |
| Scheduled | FM (D10) | (missing data) | (missing data) |
| Scheduled | FM (D10) | Presbyopia | Ophthalmology |
| Scheduled | FM (D10) | (missing data) | (missing data) |
| Scheduled | FM (D5) | Ankle sprain R | Physician licensed to treat work accidents (German: *Durchgangsarzt*) |
| Scheduled | FM (D10) | (missing data) | (missing data) |
| Scheduled | FM (D5) | Adverse drug reaction | Urology |
| Scheduled | FM (D10) | Toothache | Dentistry |
| Unscheduled | FM (D9) | Prison admission exam | Psychiatry or drug counseling (prison psychiatrist) |
| Scheduled | PSY (D8) | Chronic arm pain after car accident, insomnia[c] | Neurology |
| Scheduled | PSY (D8) | Liver cirrhosis, suspected ascites[c] | Gastroenterology |
| Scheduled | FM (D5) | Priapism | Urology |
| Scheduled | FM (D5) | First degree hemorrhoids | Proctology[d] |
| Scheduled | FM (D5) | First degree hemorrhoids | Proctology[d] |

[a]FM: family medicine.

[b]PSY: psychiatry.

[c]No diagnosis coded—information taken from medical history.

[d]Referral 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).

| Type of consultation | Method of communication | Medical specialty (physician number) | Diagnosis |
|---|---|---|---|
| Unscheduled | Video | FM[a] (D1) | Suspected tuberculosis, hypotension |
| Unscheduled | Video | FM (D7) | Unstable angina |
| Unscheduled | Video | FM (D1) | Severe hypotension suspected due to antipsychotics |
| Unscheduled | Video | FM (D11) | Gastrointestinal hemorrhage |
| Unscheduled | Phone | FM (D2) | Abdominal pain, kidney stones |
| Unscheduled | Video | FM (D2) | Gallbladder disease |
| Unscheduled | Video | FM (D4) | Abdominal pain (differential diagnosis: myocardial infarction) |
| Scheduled | Video | PSY[b] (D8) | Suspected schizophrenia |

[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.

| Prison | Video consultations per prison, (n=435), n (%) | Referrals per prison, (n=17), n (%) | Hospital admissions per prison, (n=8), n (%) |
|---|---|---|---|
| P1 | 41 (9) | 0 (0) | 2 (25) |
| P2 | 194 (45) | 5 (29) | 1 (13) |
| P3 | 41 (9) | 6 (35) | 0 (0) |
| P4 | 75 (17) | 4 (24) | 0 (0) |
| P5 | 68 (16) | 2 (12) | 4 (50) |
| P6 | 16 (4) | 0 (0) | 1 (13) |

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

| Physician | Video consultations per physician, (n=435), n (%) | Referrals per physician, (n=17), n (%) | Hospital admissions, (n=8), n (%) |
|---|---|---|---|
| D1 | 143 (33) | 1 (6) | 2 (25) |
| D2 | 9 (2) | 0 (0) | 2 (25) |
| D3 | 5 (2) | 0 (0) | 0 (0) |
| D4 | 18 (4) | 0 (0) | 1 (13) |
| D5 | 14 (3) | 6 (35) | 0 (0) |
| D6 | 8 (2) | 0 (0) | 0 (0) |
| D7 | 2 (1) | 1 (6) | 1 (13) |
| D8 | 124 (29) | 3 (17) | 1 (13) |
| D9 | 9 (2) | 2 (12) | 0 (0) |
| D10 | 97 (22) | 4 (24) | 0 (0) |
| D11 | 3 (1) | 0 (0) | 1 (13) |
| D12 | 3 (1) | 0 (0) | 0 (0) |

## 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-Württenberg. 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-Württenberg 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

1. Number of prisoners and detainees in the prisons in Germany by type of detention in June 2022. Statista. URL: https://de.statista.com/statistik/daten/studie/158317/umfrage/gefangene-und-verwahrte-in-deutschland-nach-art-des-vollzugs/ [accessed 2024-06-03]
2. Number of prisoners in the prisons in Germany after the expected execution on 31. March 2022. Statista. URL: https://de.statista.com/statistik/daten/studie/75829/umfrage/strafgefangene-nach-vollzugsdauer-in-deutschland/#statisticContainer [accessed 2024-06-03]
3. Opitz-Welke A, Lehmann M, Seidel P, Konrad N. Medicine in the penal system. Dtsch Arztebl Int 2018;115(48):808-814 [FREE Full text] [doi: 10.3238/arztebl.2018.0808] [Medline: 30642429]
4. Davis DM, Bello JK, Rottnek F. Care of incarcerated patients. Am Fam Physician 2018;98(10):577-583 [FREE Full text] [Medline: 30365288]
5. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
6. Dissemination of telemedical methods and procedures. Bundesärztkammer - Verbreitung telemedizinischer Methoden und Verfahren. URL: https://www.bundesaerztekammer.de/themen/aerzte/digitalisierung/telemedizin-fernbehandlung/verbreitung-methoden-und-verfahren [accessed 2024-06-06]
7. Mateo M, Álvarez R, Cobo C, Pallas J, López AM, Gaite L. Telemedicine: contributions, difficulties and key factors for implementation in the prison setting. Rev Esp Sanid Penit 2019;21(2):95-105 [FREE Full text] [Medline: 31642860]
8. Anogianakis G, Ilonidis G, Milliaras S, Anogeianaki A, Vlachakis-Milliaras E. Developing prison telemedicine systems: the Greek experience. J Telemed Telecare 2003;9(Suppl 2):S4-S7. [doi: 10.1258/135763303322596101] [Medline: 14728746]

9. Edge C, Black G, King E, George J, Patel S, Hayward A. Improving care quality with prison telemedicine: the effects of context and multiplicity on successful implementation and use. J Telemed Telecare 2021;27(6):325-342. [doi: 10.1177/1357633X19869131] [Medline: 31640460]
10. Giraudeau N, Inquimbert C, Delafoy R, Tramini P, Valcarcel J, Meroueh F. Teledentistry, new oral care tool for prisoners. Int J Prison Health 2017;13(2):124-134. [doi: 10.1108/IJPH-04-2016-0011] [Medline: 28581375]
11. Colombo MG, Koch R, Joos S. Akzeptanz von videokonsultationen in justizvollzugsanstalten aus der perspektive von patient/-innen—ergebnisse der gemischt-methodischen evaluation eines modellprojekts aus baden-württemberg [Acceptance of video consultations in correctional facilities from the patients' perspective—results from the mixed-methods evaluation of a pilot project in baden-württemberg]. Psychiatr Prax 2022;49(2):80-88. [doi: 10.1055/a-1400-2388] [Medline: 33773499]
12. Ellis DG, Mayrose J, Jehle DV, Moscati RM, Pierluisi GJ. A telemedicine model for emergency care in a short-term correctional facility. Telemed J E Health 2001;7(2):87-92. [doi: 10.1089/153056201750279584] [Medline: 11421081]
13. Fox KC, Somes GW, Waters TM. Timeliness and access to healthcare services via telemedicine for adolescents in state correctional facilities. J Adolesc Health 2007;41(2):161-167. [doi: 10.1016/j.jadohealth.2007.05.001] [Medline: 17659220]
14. Fox KC, Somes G, Waters TM. The use of telemedicine technology and restructured referral patterns to reduce health care costs in a Juvenile Justice System. J Correct Health Care 2006;12(3):214-221. [doi: 10.1177/1078345806292667]
15. Kassar K, Roe C, Desimone M. Use of telemedicine for management of diabetes in correctional facilities. Telemed J E Health 2017;23(1):55-59. [doi: 10.1089/tmj.2016.0036] [Medline: 27223479]
16. Galán GJ, Alia CA, González MV, Berriguete R, González FF, Rodríguez CMF, et al. The contribution of telemedicine to hepatitis C elimination in a correctional facility. Rev Esp Enferm Dig 2019;111(7):550-555 [FREE Full text] [doi: 10.17235/reed.2019.6152/2018] [Medline: 31215210]
17. Holland A. Telehealth reduces hospital admission rates in patients with COPD. J Physiother 2013;59(2):129 [FREE Full text] [doi: 10.1016/S1836-9553(13)70168-1] [Medline: 23663801]
18. Cruz J, Brooks D, Marques A. Home telemonitoring effectiveness in COPD: a systematic review. Int J Clin Pract 2014;68(3):369-378. [doi: 10.1111/ijcp.12345] [Medline: 24472009]
19. Rosen D, McCall JD, Primack BA. Telehealth protocol to prevent readmission among high-risk patients with congestive heart failure. Am J Med 2017;130(11):1326-1330. [doi: 10.1016/j.amjmed.2017.07.007] [Medline: 28756266]
20. O'Connor M, Asdornwised U, Dempsey ML, Huffenberger A, Jost S, Flynn D, et al. Using telehealth to reduce all-cause 30-day hospital readmissions among heart failure patients receiving skilled home health services. Appl Clin Inform 2016;7(2):238-247 [FREE Full text] [doi: 10.4338/ACI-2015-11-SOA-0157] [Medline: 27437037]
21. Joseph JW, Kennedy M, Nathanson LA, Wardlow L, Crowley C, Stuck A. Reducing emergency department transfers from skilled nursing facilities through an emergency physician telemedicine service. West J Emerg Med 2020;21(6):205-209 [FREE Full text] [doi: 10.5811/westjem.2020.7.46295] [Medline: 33207167]
22. Rosner BI, Gottlieb M, Anderson WN. Effectiveness of an automated digital remote guidance and telemonitoring platform on costs, readmissions, and complications after hip and knee arthroplasties. J Arthroplasty 2018;33(4):988-996.e4 [FREE Full text] [doi: 10.1016/j.arth.2017.11.036] [Medline: 29229238]
23. Huang VW, Reich KM, Fedorak RN. Distance management of inflammatory bowel disease: systematic review and meta-analysis. World J Gastroenterol 2014;20(3):829-842 [FREE Full text] [doi: 10.3748/wjg.v20.i3.829] [Medline: 24574756]
24. Liddy C, Bello A, Cook J, Drimer N, Pilon MD, Farrell G, et al. Supporting the spread and scale-up of electronic consultation across Canada: cross-sectional analysis. BMJ Open 2019;9(5):e028888 [FREE Full text] [doi: 10.1136/bmjopen-2018-028888] [Medline: 31152043]
25. Liddy C, Moroz I, Keely E, Taljaard M, Fraser AM, Armstrong CD, et al. The use of electronic consultations is associated with lower specialist referral rates: a cross-sectional study using population-based health administrative data. Fam Pract 2018;35(6):698-705 [FREE Full text] [doi: 10.1093/fampra/cmy020] [Medline: 29635449]
26. Liddy C, Moroz I, Keely E, Taljaard M, Armstrong CD, Afkham A, et al. Understanding the impact of a multispecialty electronic consultation service on family physician referral rates to specialists: a randomized controlled trial using health administrative data. Trials 2019;20(1):348 [FREE Full text] [doi: 10.1186/s13063-019-3393-5] [Medline: 31182123]
27. Waldura JF, Neff S, Dehlendorf C, Goldschmidt RH. Teleconsultation improves primary care clinicians' confidence about caring for HIV. J Gen Intern Med 2013;28(6):793-800 [FREE Full text] [doi: 10.1007/s11606-013-2332-5] [Medline: 23371417]
28. van Gelder VA, Scherpbier-de Haan ND, van Berkel S, Akkermans RP, de Grauw IS, Adang EM, et al. Web-based consultation between general practitioners and nephrologists: a cluster randomized controlled trial. Fam Pract 2017;34(4):430-436 [FREE Full text] [doi: 10.1093/fampra/cmw131] [Medline: 28158576]
29. Ayadi HE, Desai A, Jones RE, Mercado E, Williams M, Rooks B, et al. Referral rates vary widely between family medicine practices. J Am Board Fam Med 2021;34(6):1183-1188 [FREE Full text] [doi: 10.3122/jabfm.2021.06.210213] [Medline: 34772773]
30. Crump WJ, Massengill P. Outpatient consultations from a family practice residency program: nine years' experience. J Am Board Fam Pract 1988;1(3):164-166. [Medline: 3177014]

31. Tabenkin H, Oren B, Steinmetz D, Tamir A, Kitai E. Referrals of patients by family physicians to consultants: a survey of the Israeli family practice research network. Fam Pract 1998;15(2):158-164 [FREE Full text] [doi: 10.1093/fampra/15.2.158] [Medline: 9613485]
32. Franks P, Zwanziger J, Mooney C, Sorbero M. Variations in primary care physician referral rates. Health Serv Res 1999;34(1 Pt 2):323-329 [FREE Full text] [Medline: 10199678]
33. Gwynne M, Page C, Reid A, Donahue K, Newton W. What's the right referral rate? specialty referral patterns and curricula across I3 collaborative primary care residencies. Fam Med 2017;49(2):91-96 [FREE Full text] [Medline: 28218933]
34. Duffy R, Neville R, Staines H. Variance in practice emergency medical admission rates: can it be explained? Br J Gen Pract 2002;52(474):14-17 [FREE Full text] [Medline: 11791810]
35. Blatchford O, Capewell S, Murray S, Blatchford M. Emergency medical admissions in Glasgow: general practices vary despite adjustment for age, sex, and deprivation. Br J Gen Pract 1999;49(444):551-554 [FREE Full text] [Medline: 10621990]
36. Swift C, Cain SM, Needham M. A primary care telehealth experience in a US army correctional facility in Germany. US Army Med Dep J 2016:76-80. [Medline: 26874102]
37. Brunicardi BO. Financial analysis of savings from telemedicine in Ohio's prison system. Telemed J 1998;4(1):49-54. [doi: 10.1089/tmj.1.1998.4.49] [Medline: 9599074]
38. Sullivan DH, Chapman M, Mullen PE. Videoconferencing and forensic mental health in Australia. Behav Sci Law 2008;26(3):323-331. [doi: 10.1002/bsl.815] [Medline: 18548513]
39. Deslich SA, Thistlethwaite T, Coustasse A. Telepsychiatry in correctional facilities: using technology to improve access and decrease costs of mental health care in underserved populations. Perm J 2013;17(3):80-86 [FREE Full text] [doi: 10.7812/TPP/12-123] [Medline: 24355894]
40. Aoki N, Dunn K, Fukui T, Beck JR, Schull WJ, Li HK. Cost-effectiveness analysis of telemedicine to evaluate diabetic retinopathy in a prison population. Diabetes Care 2004;27(5):1095-1101 [FREE Full text] [doi: 10.2337/diacare.27.5.1095] [Medline: 15111527]
41. Yogesan K, Henderson C, Barry CJ, Constable IJ. Online eye care in prisons in Western Australia. J Telemed Telecare 2001;7(Suppl 2):63-64. [doi: 10.1258/1357633011937173] [Medline: 11747663]
42. McCue MJ, Hampton CL, Malloy W, Fisk KJ, Dixon L, Neece A. Financial analysis of telecardiology used in a correctional setting. Telemed J E Health 2000;6(4):385-391. [doi: 10.1089/15305620050503852] [Medline: 11242546]
43. de Almeida Castro Morosini I, de Oliveira DC, de Morais Ferreira F, Fraiz FC, Torres-Pereira CC. Performance of distant diagnosis of dental caries by teledentistry in juvenile offenders. Telemed J E Health 2014;20(6):584-589. [doi: 10.1089/tmj.2013.0202] [Medline: 24693859]
44. Edge C, George J, Black G, Gallagher M, Ala A, Patel S, et al. Using telemedicine to improve access, cost and quality of secondary care for people in prison in England: a hybrid type 2 implementation effectiveness study. BMJ Open 2020;10(2):e035837 [FREE Full text] [doi: 10.1136/bmjopen-2019-035837] [Medline: 32075846]
45. Senanayake B, Wickramasinghe SI, Eriksson L, Smith AC, Edirippulige S. Telemedicine in the correctional setting: a scoping review. J Telemed Telecare 2018;24(10):669-675 [FREE Full text] [doi: 10.1177/1357633X18800858] [Medline: 30343655]
46. Brunetti ND, Dellegrottaglie G, Di Giuseppe G, De Gennaro L, Di Biase M. Prison break: remote tele-cardiology support for cardiology emergency in Italian penitentiaries. Int J Cardiol 2013;168(3):3138-3140. [doi: 10.1016/j.ijcard.2013.04.022] [Medline: 23623349]


## 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.*

*Please cite as:*
*Schmalstieg-Bahr K, Colombo MG, Koch R, Szecsenyi J, Völker F, Blozik EE, Scherer M*
*Intramural Health Care Through Video Consultations and the Need for Referrals and Hospital Admissions: Retrospective Quantitative Subanalysis of an Evaluation Study*
*Interact J Med Res 2024;13:e44906*
*URL: https://www.i-jmr.org/2024/1/e44906*
*doi:10.2196/44906*
*PMID:*

Original Paper

# Evolution of Digital Health and Exploration of Patented Technologies (2017-2021): Bibliometric Analysis

Wenjun Gu[1], MPH; Jinhua Wang[2], BA; Yunqi Zhang[3], MSc; Shaolin Liang[3,4*], PhD; Zisheng Ai[1*], PhD; Jiyu Li[5*], MD

[1]School of Medicine, Tongji University, Shanghai, China

[2]Shanghai Institute of Medical Innovation Technology Transfer, Shanghai, China

[3]STI-Zhilian Research Institute for Innovation and Digital Health, Beijing, China

[4]Institute for Six-sector Economy, Fudan University, Shanghai, China

[5]Geriatric 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.

***(Interact J Med Res 2024;13:e48259)*** doi:10.2196/48259

**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.

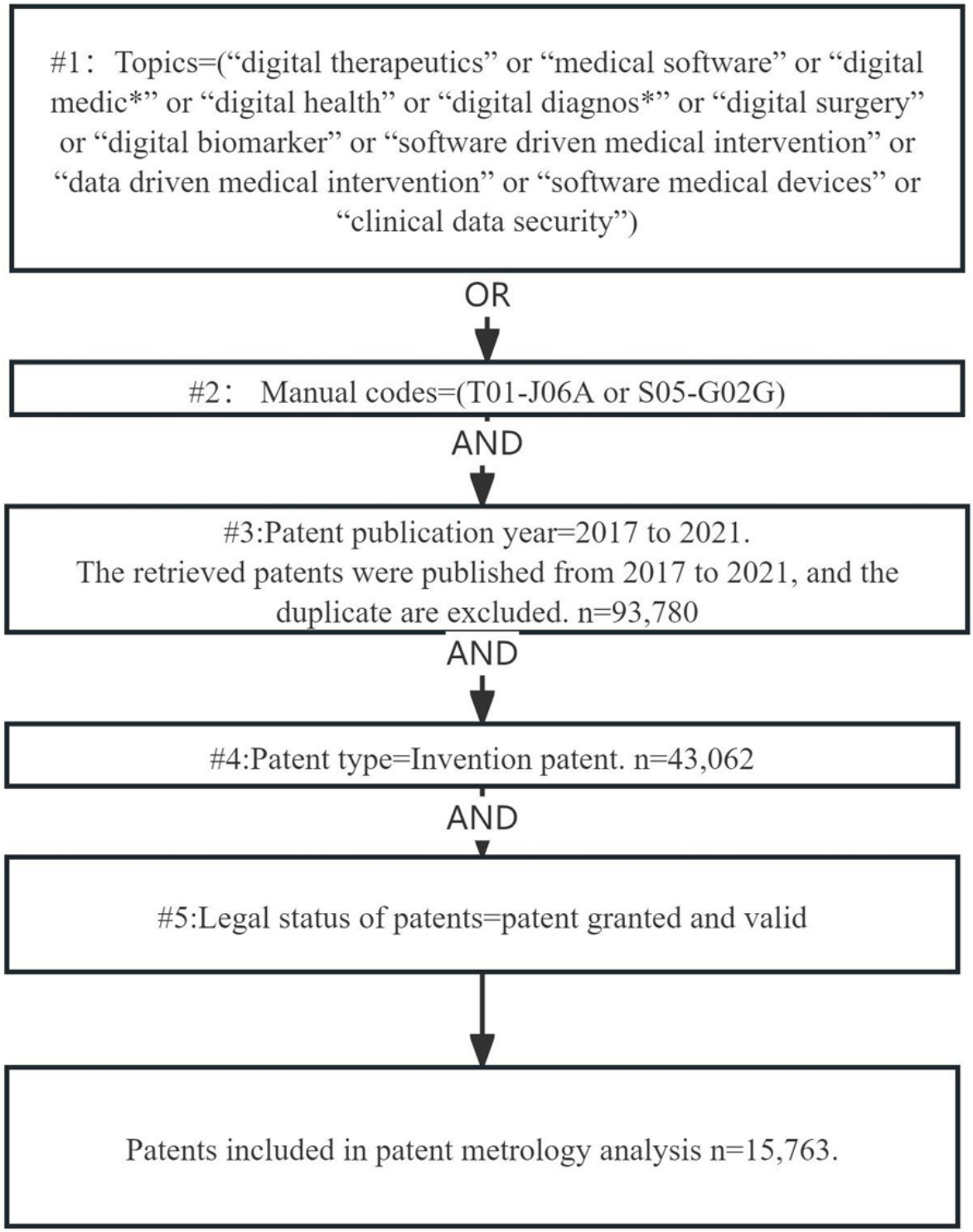
## Methods

### 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 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.

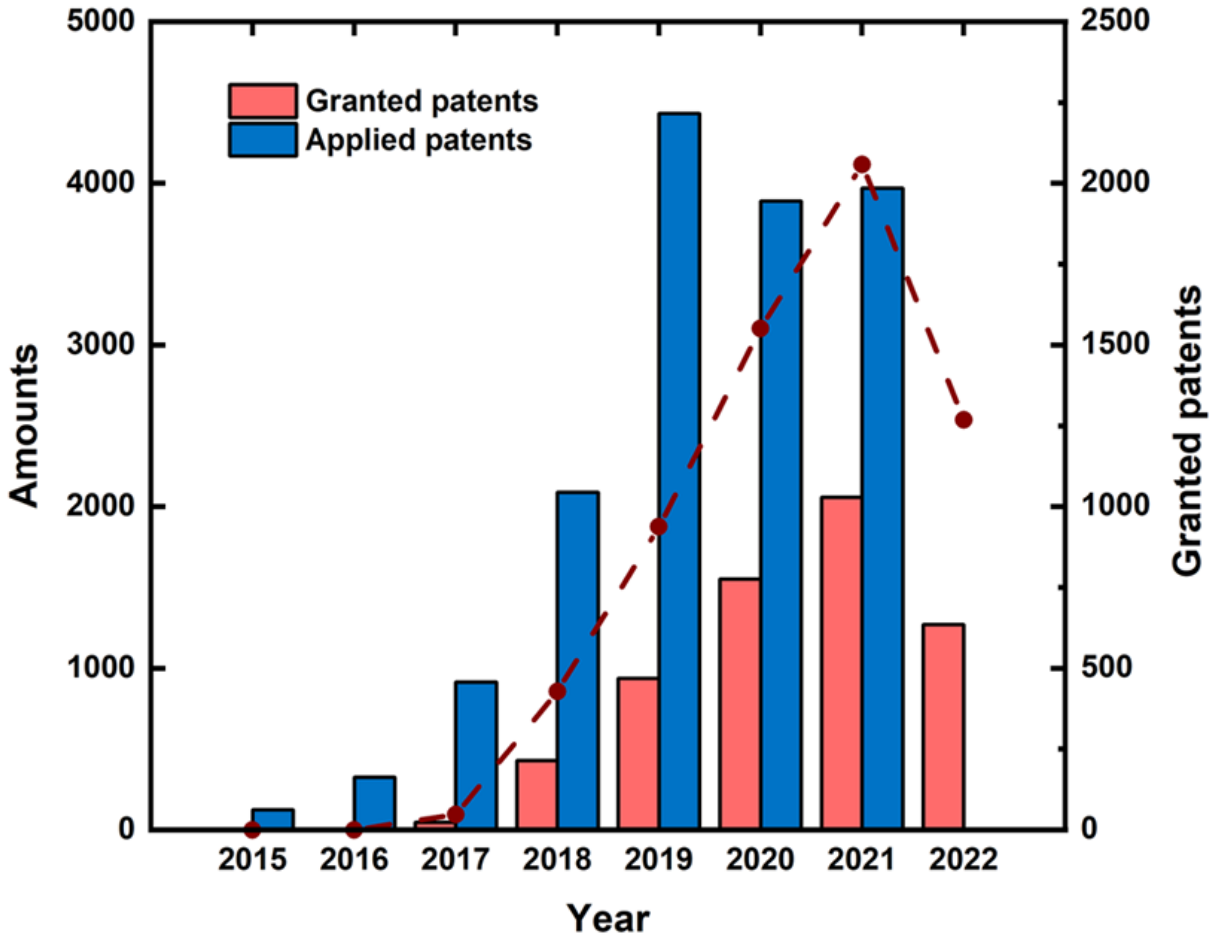
The chosen records were exported from the Derwent Innovation Index in a table format and subsequently imported into Excel 2016 for thorough analysis. The study focused on examining the annual volume of applications, granted patents, the most productive countries, institutions, inventors, highly cited 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.

**Figure 2.** Applied and granted patents from 2017 to 2022.



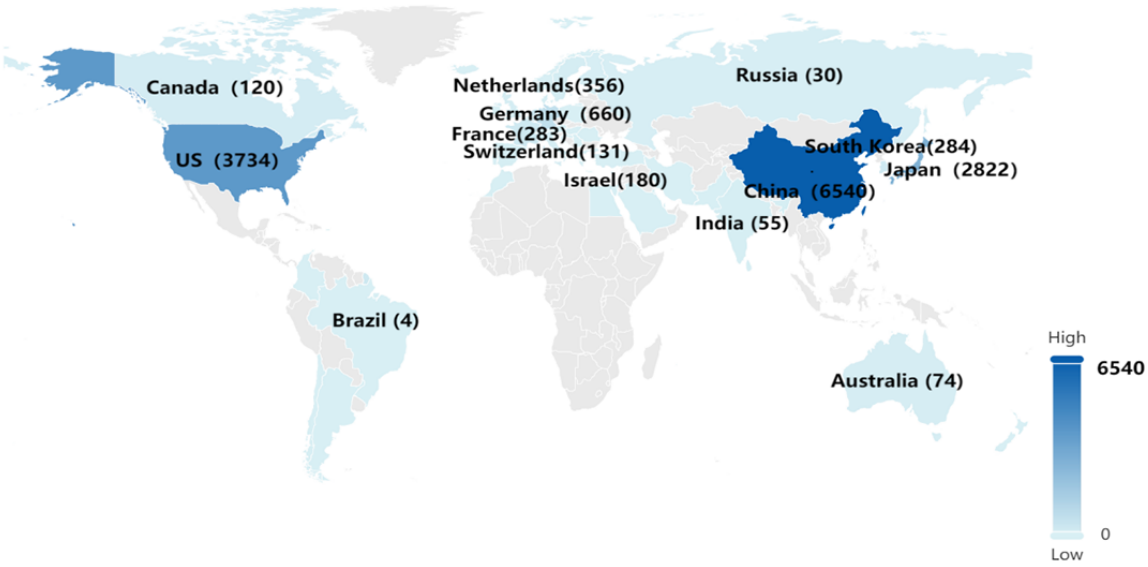

## Geographical Distribution

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).

| Intellectual property office | Published patents, n (%) |
|---|---|
| China National Intellectual Property Administration | 7253 (46.01) |
| United States Patent and Trademark Office | 4052 (25.71) |
| Japanese patent office | 2174 (13.79) |
| European patent office | 1242 (7.88) |
| IP Australia | 244 (1.55) |
| Inspirations Property Institution | 228 (1.45) |
| World Intellectual Property Organization | 216 (1.37) |
| Deutsches Patent- und Markenamt | 172 (1.09) |
| Canadian Intellectual Property Office | 94 (0.6) |
| Spanish Patent and Trademark Office | 46 (0.29) |

**Figure 3.** Intellectual property office for published patents (marked according to the country where the intellectual property office is located).




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).

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

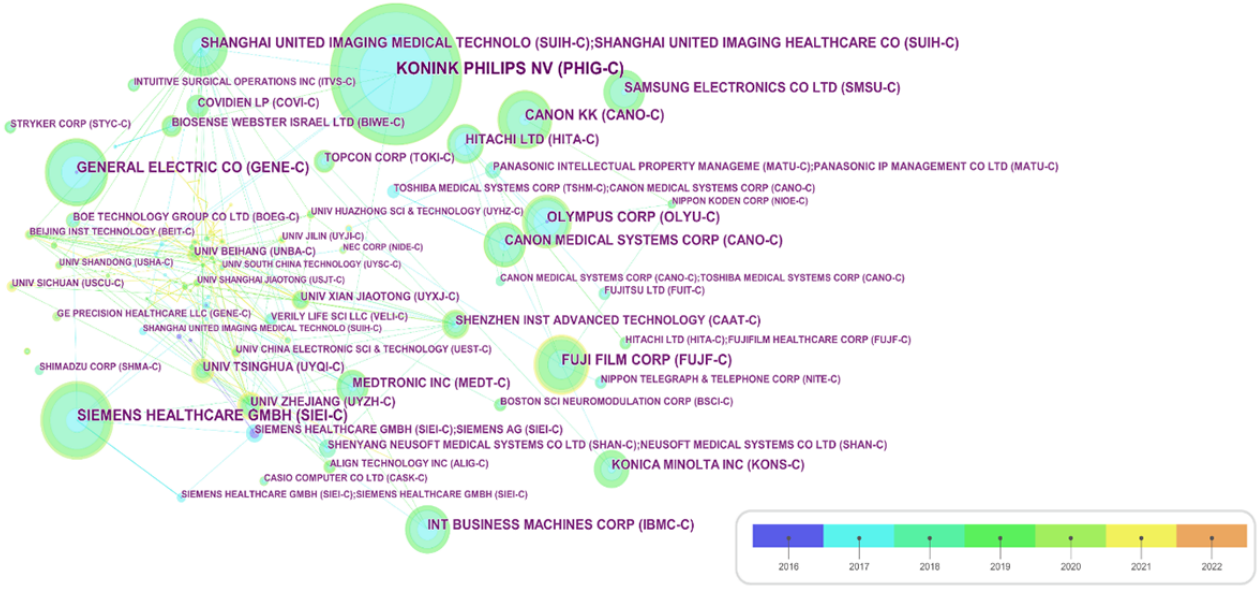
Table 3 highlights the top 10 highly cited institutions, all of which are enterprises. Auris Health Inc and Masimo Corporation stand out with 849 and 983 citations globally, respectively, underscoring their substantial technological influence in the digital health field. It is noteworthy that Lepu (Beijing)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

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

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

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.

**Figure 4.** Institutional coauthorship network.




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 15763. 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.

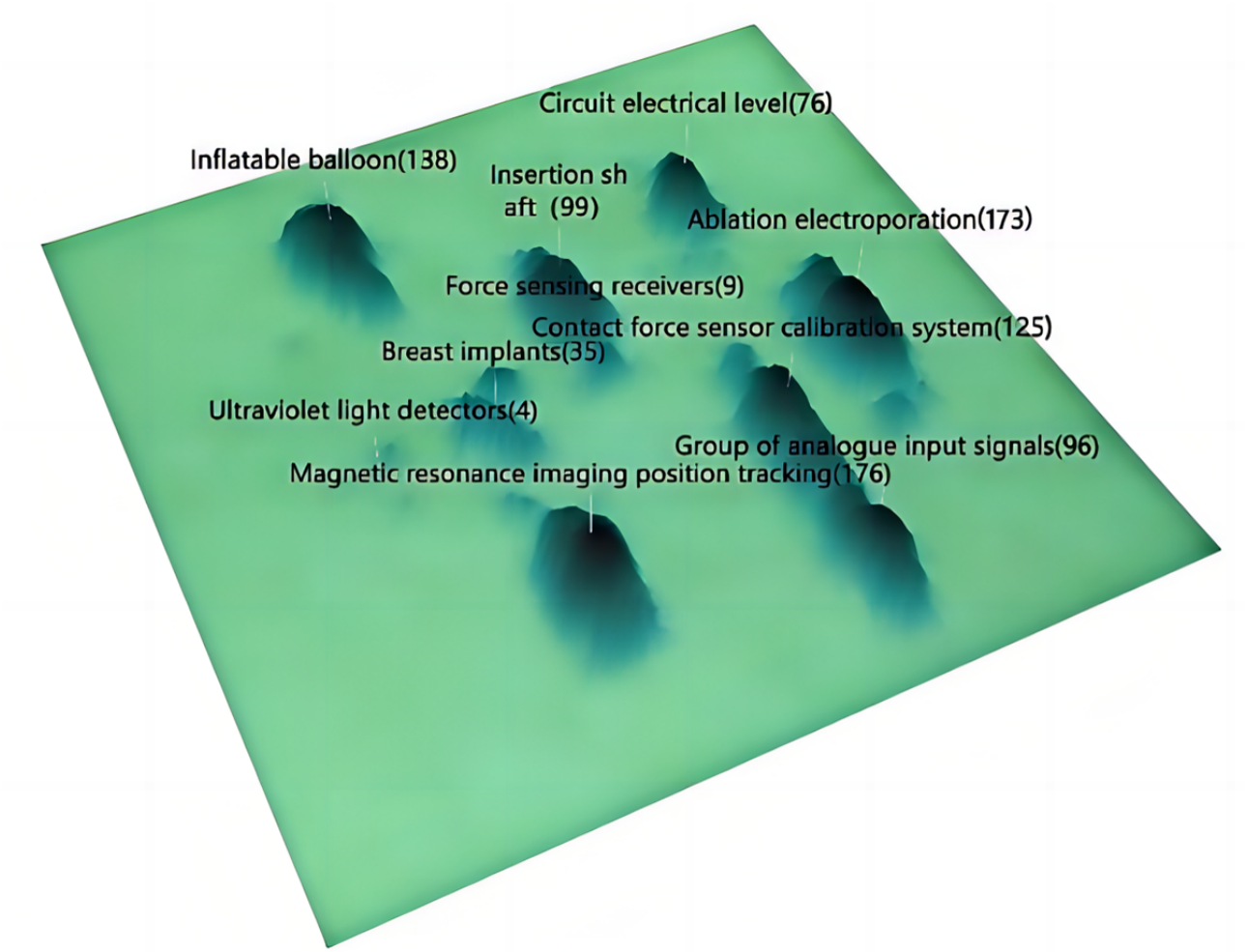
**Table 4.** Top 10 inventors with published patents.

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

[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.

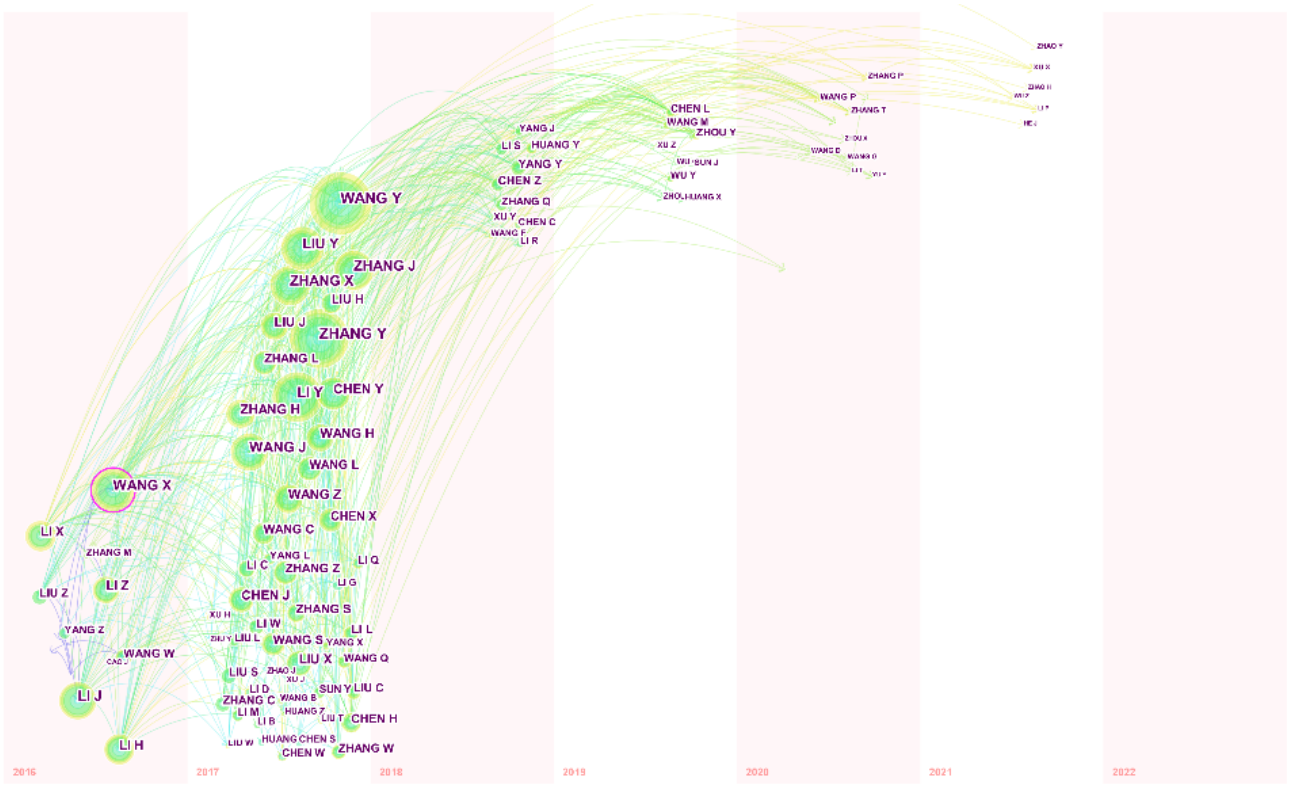



Ammar Al Ali held a position as a fellow at Masimo Corporation, a global health care technology enterprise renowned for its research and development of cutting-edge noninvasive patient monitoring techniques. Masimo Corporation also secured the second position among the top 10 leading institutes globally, 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.

| Content | | Published patents, n |
|---|---|---|
| **Derwent Manual Code** | | |
| T01-J06A | Medical equipment and information systems | 14,579 |
| P31-A05 | Diagnostic devices | 9081 |
| T01-S03 | Claimed software products | 3781 |
| T01-J10B2 | Image analysis | 2573 |
| S05-D | Electrical diagnosis | 1982 |
| B11-C11 | General computing methods and apparatus | 1278 |
| S05-D01 | Measuring and recording systems | 804 |
| T01-J10B1 | Image enhancement | 1028 |
| P34-A02 | Syringes for removing and introducing fluids into the body | 933 |
| P31-A01 | Surgical tools and instruments | 894 |
| S05-D07 | Diagnostic displays and monitors | 756 |
| **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[a] dedicated to arranging or managing health care resources or facilities | 659 |
| G16H40 | ICT dedicated to operating medical equipment or devices | 650 |

[a]ICT: 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.

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

[a]MRI: magnetic resonance imaging.

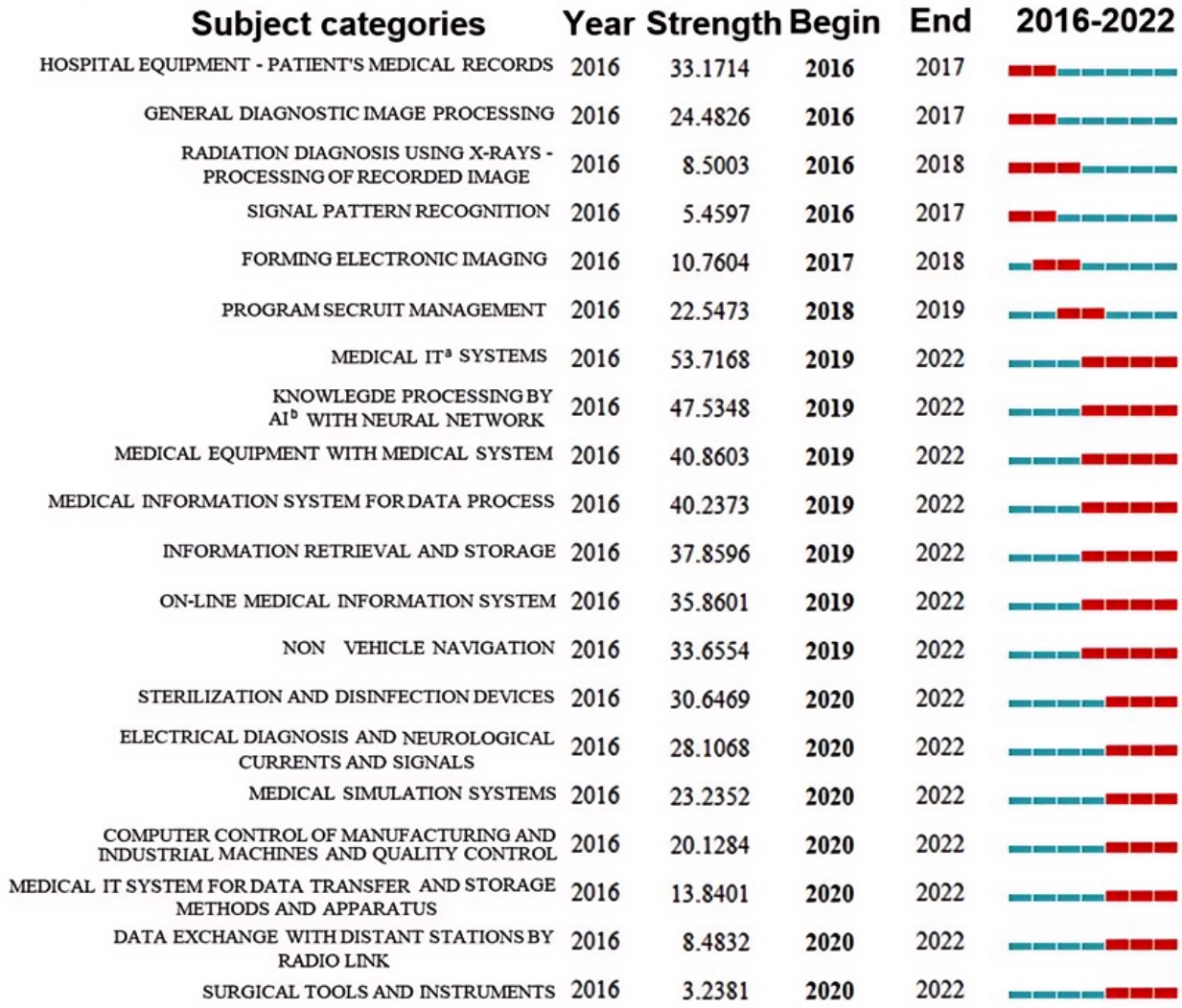
[b]CT: computed tomography.

[c]AI: artificial intelligence.

[d]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.

**Figure 7.** Top 20 subject categories with the strongest citation bursts.




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," "web-based medical information system," "medical simulation systems," "medical information technology system for data transfer or storage," and "medical information system for data processing." The third category encompasses methods for processing medical data, including themes such as "signal pattern recognition," "forming electronic image," "knowledge processing by artificial intelligence with neural networks," "information retrieval and storage," and "data exchange with distant stations by radio link." The last category pertains to the specific application of digital health technology in medical fields, 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 development, production, and sales of health care products, 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, medical, and military 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 road map 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

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
Nationality distribution of patent applicants within the China National Intellectual Property Administration, Japanese Patent Office, and United States Patent and Trademark Office.
[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


1. Thompson M. The geographies of digital health - digital therapeutic landscapes and mobilities. Health Place 2021 Jul;70:102610. [doi: 10.1016/j.healthplace.2021.102610] [Medline: 34174771]
2. Digital therapeutics alliance homepage. Digital Therapeutics Alliance. URL: https://dtxalliance.org/ [accessed 2021-08-09]
3. Digital therapeutics definition and core principles. Digital Therapeutics Alliance. URL: https://dtxalliance.org/wp-content/uploads/2021/01/DTA_DTx-Definition-and-Core-Principles.pdf [accessed 2021-09-10]
4. Ramakrishnan P, Yan K, Balijepalli C, Druyts E. Changing face of healthcare: digital therapeutics in the management of diabetes. Curr Med Res Opin 2021 Dec;37(12):2089-2091. [doi: 10.1080/03007995.2021.1976737] [Medline: 34511002]
5. Canonico ME, Hsia J, Guthrie NL, Simmons M, Mehta P, Lupinacci P, et al. Cognitive behavioral therapy delivered via digital mobile application for the treatment of type 2 diabetes: rationale, design, and baseline characteristics of a randomized, controlled trial. Clin Cardiol 2022 Aug;45(8):850-856 [FREE Full text] [doi: 10.1002/clc.23853] [Medline: 35778834]
6. Hildebrand AS, Roesmann K, Planert J, Machulska A, Otto E, Klucken T. Self-guided virtual reality therapy for social anxiety disorder: a study protocol for a randomized controlled trial. Trials 2022 May 12;23(1):395 [FREE Full text] [doi: 10.1186/s13063-022-06320-x] [Medline: 35549980]
7. Abbadessa G, Brigo F, Clerico M, De Mercanti S, Trojsi F, Tedeschi G, et al. Digital therapeutics in neurology. J Neurol 2022 Mar 20;269(3):1209-1224 [FREE Full text] [doi: 10.1007/s00415-021-10608-4] [Medline: 34018047]
8. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan 29;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
9. Guthrie NL, Carpenter J, Edwards KL, Appelbaum KJ, Dey S, Eisenberg DM, et al. Emergence of digital biomarkers to predict and modify treatment efficacy: machine learning study. BMJ Open 2019 Jul 23;9(7):e030710 [FREE Full text] [doi: 10.1136/bmjopen-2019-030710] [Medline: 31337662]
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]
11. Nielsen P, Sahay S. A critical review of the role of technology and context in digital health research. Digit Health 2022 Jun 22;8:20552076221109554 [FREE Full text] [doi: 10.1177/20552076221109554] [Medline: 35769359]
12. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022 Jul 06;24(7):e35816 [FREE Full text] [doi: 10.2196/35816] [Medline: 35793141]
13. Ahmadvand A, Kavanagh D, Clark M, Drennan J, Nissen L. Trends and visibility of "digital health" as a keyword in articles by JMIR publications in the new millennium: bibliographic-bibliometric analysis. J Med Internet Res 2019 Dec 19;21(12):e10477 [FREE Full text] [doi: 10.2196/10477] [Medline: 31855190]
14. Gupta BM, Dhawan SM, Mueen Ahmed KK. Digital health research: a scientometric assessment of global publications output during 2007–2016. Int J Pharm Investig 2018 Sep;8(2) [FREE Full text] [doi: 10.4103/jphi.JPHI_31_18]
15. Sweileh WM, Al-Jabi SW, AbuTaha AS, Zyoud SH, Anayah FM, Sawalha AF. Bibliometric analysis of worldwide scientific literature in mobile - health: 2006-2016. BMC Med Inform Decis Mak 2017 May 30;17(1):72 [FREE Full text] [doi: 10.1186/s12911-017-0476-7] [Medline: 28558687]
16. Yao L, Hui L, Yang Z, Chen X, Xiao A. Freshwater microplastics pollution: detecting and visualizing emerging trends based on Citespace II. Chemosphere 2020 Apr;245:125627. [doi: 10.1016/j.chemosphere.2019.125627] [Medline: 31864046]
17. Chen C. Science mapping: a systematic review of the literature. J Data Inf Sci 2017 Feb;2(2):1-40. [doi: 10.1515/jdis-2017-0006]
18. Derwent World Patents Index (DWPI): GMPI and EPI manual codes (part 1). Clarivate Analytics. URL: https://clarivate.com/derwent/wp-content/uploads/sites/3/dlm_uploads/2019/08/EPI-Manual-Codes-Part-1-2020.pdf [accessed 2022-09-10]
19. DWPI classification system. Clarivate. URL: https://clarivate.com/dwpi-reference-center/dwpi-classification-system/ [accessed 2022-09-10]
20. Chen C, Hu Z, Liu S, Tseng H. Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. Expert Opin Biol Ther 2012 May;12(5):593-608. [doi: 10.1517/14712598.2012.674507] [Medline: 22443895]
21. Wang C, Wu J, He X, Ye M, Liu W, Tang R. Emerging trends and new developments in disaster research after the 2008 Wenchuan earthquake. Int J Environ Res Public Health 2018 Dec 23;16(1):29 [FREE Full text] [doi: 10.3390/ijerph16010029] [Medline: 30583598]
22. Wang L, Xue X, Zhang Y, Luo X. Exploring the emerging evolution trends of urban resilience research by scientometric analysis. Int J Environ Res Public Health 2018 Oct 06;15(10):2181 [FREE Full text] [doi: 10.3390/ijerph15102181] [Medline: 30301209]
23. Li X, Ma E, Qu H. Knowledge mapping of hospitality research − a visual analysis using CiteSpace. Int J Hosp Manag 2017 Jan;60:77-93. [doi: 10.1016/j.ijhm.2016.10.006]
24. Liang YD, Li Y, Zhao J, Wang XY, Zhu HZ, Chen XH. Study of acupuncture for low back pain in recent 20 years: a bibliometric analysis via CiteSpace. J Pain Res 2017 Apr 24;10:951-964 [FREE Full text] [doi: 10.2147/JPR.S132808] [Medline: 28479858]
25. Fang Y, Yin J, Wu B. Climate change and tourism: a scientometric analysis using CiteSpace. J Sustain Tour 2017 May 30;26(1):108-126. [doi: 10.1080/09669582.2017.1329310]

26. Chen X, Li F, Zhang J, Zhou W, Wang X, Fu H. Spatiotemporal mapping and multiple driving forces identifying of PM2.5 variation and its joint management strategies across China. J Clean Prod 2020 Mar 20;250:119534. [doi: 10.1016/j.jclepro.2019.119534]
27. Intelligent data analytics software solution comes to electrophysiology with launch of Biosense Webster's Cartonet. Johnson & Johnson MedTech. 2019 Jul 22. URL: https://www.jnjmedtech.com/en-US/news-events/biosense-webster-launches-cartonet [accessed 2024-06-12]
28. Inventing the future: an introduction to patents for small and medium-sized enterprises. World Intellectual Property Organization. URL: https://www.wipo.int/edocs/pubdocs/en/sme/917/wipo_pub_917.pdf [accessed 2022-09-10]
29. Systems and methods for automated detection in magnetic resonance images. Google Patents. URL: https://patents.google.com/patent/EP3545494A4/en [accessed 2024-06-12]
30. Enhanced convolutional neural network for image segmentation. Google Patents. 2018. URL: https://patents.google.com/patent/US10140544B1/en?oq=US15943392 [accessed 2024-06-17]
31. Systems and methods for medical instrument compression compensation. Google Patents. URL: https://patents.google.com/patent/JP2022177148A/en [accessed 2024-06-12]
32. Tran B, Tran H. Systems and methods for computer assisted operation. Justia Patents. 2019. URL: https://patents.justia.com/inventor/ha-tran?page=6 [accessed 2024-06-12]
33. Robotic system with indication of boundary for robotic arm. Auris Health, Inc. 2019 Apr 11. URL: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2019070482 [accessed 2024-06-12]
34. Detection of misalignment of robotic arms. Auris Health, Inc. URL: https://portal.unifiedpatents.com/patents/patent/US-D932628-S1 [accessed 2024-06-12]
35. Authentication and information system for reusable surgical instruments. Justia Patents. 2019. URL: https://patents.justia.com/patent/10891149 [accessed 2024-06-12]
36. Motorized surgical instrument with an end effector. Google Patents. URL: https://patents.google.com/patent/US10980535B2/en [accessed 2024-06-12]
37. Detecting and using body tissue electrical signals. Google Patents. URL: https://patents.google.com/patent/US20180153430A1/en [accessed 2024-06-12]
38. Opioid overdose monitoring. Google Patents. URL: https://patents.google.com/patent/US10932729B2/en [accessed 2024-06-12]
39. Sikandar H, Abbas AF, Khan N, Qureshi MI. Digital technologies in healthcare: a systematic review and bibliometric analysis. Int J Online Biomed Eng 2022;18(8):34-48. [doi: 10.3991/ijoe.v18i08.31961]


## Abbreviations

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


*Edited by T de Azevedo Cardoso; submitted 18.04.23; peer-reviewed by Y Ma, W Tumuhimbise; comments to author 30.11.23; revised version received 24.01.24; accepted 18.04.24; published 11.07.24.*

*Please cite as:*
*Gu W, Wang J, Zhang Y, Liang S, Ai Z, Li J*
*Evolution of Digital Health and Exploration of Patented Technologies (2017-2021): Bibliometric Analysis*
*Interact J Med Res 2024;13:e48259*
*URL: https://www.i-jmr.org/2024/1/e48259*
*doi:10.2196/48259*
*PMID:*

Original Paper

# Emerging Indications for Hyperbaric Oxygen Treatment: Registry Cohort Study

Hideaki L Tanaka[1], MD; Judy R Rees[2], BM, BChD, MPH; Ziyin Zhang[2], MS; Judy A Ptak[3], RN; Pamela M Hannigan[3], RN; Elaine M Silverman[3], MD; Janet L Peacock[2], MSc, PhD; Jay C Buckey[2], MD; Multicenter Registry for Hyperbaric Oxygen Treatment Consortium[4]

[1]Division of Undersea and Hyperbaric Medicine, Department of Emergency Medicine, University of California at San Diego, San Diego, CA, United States

[2]Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States

[3]Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States

[4]see Acknowledgments

**Corresponding Author:**
Jay C Buckey, MD
Geisel School of Medicine
Dartmouth College
1 Medical Center Drive
Lebanon, NH, 03756
United States
Phone: 1 603 646 5328
Fax: 1 603 646 5328
Email: jay.buckey@dartmouth.edu

## Abstract

**Background:** Hyperbaric oxygen ($HBO_2$) treatment is used across a range of medical specialties for a variety of applications, particularly where hypoxia and inflammation are important contributors. Because of its hypoxia-relieving and anti-inflammatory effects $HBO_2$ may be useful for new indications not currently approved by the Undersea and Hyperbaric Medical Society. Identifying these new applications for $HBO_2$ is difficult because individual centers may only treat a few cases and not track the outcomes consistently. The web-based International Multicenter Registry for Hyperbaric Oxygen Therapy captures prospective outcome data for patients treated with $HBO_2$ therapy. These data can then be used to identify new potential applications for $HBO_2$, which has relevance for a range of medical specialties.

**Objective:** Although hyperbaric medicine has established indications, new ones continue to emerge. One objective of this registry study was to identify cases where $HBO_2$ has been used for conditions falling outside of current Undersea and Hyperbaric Medical Society–approved indications and present outcome data for them.

**Methods:** This descriptive study used data from a web-based, multicenter, international registry of patients treated with $HBO_2$. Participating centers agree to collect data on all patients treated using standard outcome measures, and individual centers send deidentified data to the central registry. $HBO_2$ treatment programs in the United States, the United Kingdom, and Australia participate. Demographic, outcome, complication, and treatment data, including pre- and posttreatment quality of life questionnaires (EQ-5D-5L) were collected for individuals referred for $HBO_2$ treatment.

**Results:** Out of 9726 patient entries, 378 (3.89%) individuals were treated for 45 emerging indications. Post–COVID-19 condition (PCC; also known as postacute sequelae of COVID-19; 149/378, 39.4%), ulcerative colitis (47/378, 12.4%), and Crohn disease (40/378, 10.6%) accounted for 62.4% (n=236) of the total cases. Calciphylaxis (20/378, 5.3%), frostbite (18/378, 4.8%), and peripheral vascular disease–related wounds (12/378, 3.2%) accounted for a further 13.2% (n=50). Patients with PCC reported significant improvement on the Neurobehavioral Symptom Inventory (NSI score: pretreatment=30.6; posttreatment=14.4; $P$<.001). Patients with Crohn disease reported significantly improved quality of life (EQ-5D score: pretreatment=53.8; posttreatment=68.8), and 5 (13%) reported closing a fistula. Patients with ulcerative colitis and complete pre- and post-$HBO_2$ data reported improved quality of life and lower scores on a bowel questionnaire examining frequency, blood, pain, and urgency. A subset of patients with calciphylaxis and arterial ulcers also reported improvement.

**Conclusions:** $HBO_2$ is being used for a wide range of possible applications across various medical specialties for its hypoxia-relieving and anti-inflammatory effects. Results show statistically significant improvements in patient-reported outcomes for inflammatory bowel disease and PCC. $HBO_2$ is also being used for frostbite, pyoderma gangrenosum, pterygium, hypospadias repair, and facial filler procedures. Other indications show evidence for improvement, and the case series for all indications is growing in the registry.

**International Registered Report Identifier (IRRID):** RR2-10.2196/18857

***(Interact J Med Res 2024;13:e53821)*** doi:10.2196/53821

**KEYWORDS**

hyperbaric oxygen; inflammatory bowel disease; calciphylaxis; post–COVID-19 condition; PCC; postacute sequelae of COVID-19; PASC; infected implanted hardware; hypospadias; frostbite; facial filler; pyoderma gangrenosum

## Introduction

### Background

Hypoxia and inflammation are part of the pathophysiology for various conditions across a range of medical subspecialties. One approach to relieving hypoxia and reducing inflammation is the use of hyperbaric oxygen ($HBO_2$). $HBO_2$ is 100% oxygen delivered at pressures >1.4 atmospheres absolute (ATA) within a pressurized chamber. Typically, pressures of ≥2.0 ATA are used. $HBO_2$ treatments greatly increase the amount of oxygen dissolved in plasma and tissue during the treatment and are very effective for relieving hypoxia. The high levels of oxygen in tissue lead to a variety of biochemical effects including reduced inflammation and the release of stem cells from the bone marrow [1,2], leading to its application in several conditions, often in cases when standard treatment is not effective.

Currently, the Undersea and Hyperbaric Medical Society (UHMS) has identified 15 conditions where $HBO_2$ can be considered an approved treatment [3]. These range from caisson disease (decompression illness), where the combination of increased pressure, relief of hypoxia, and reduced inflammation from $HBO_2$ help combat the impaired circulation and endothelial damage caused by bubbles [4], to radiation injury, where the pulses of oxygen promote angiogenesis and wound healing [3]. Because of $HBO_2$'s effects on hypoxia and inflammation, more medical diagnoses likely exist that can benefit from $HBO_2$. However, $HBO_2$ is typically given in long courses (20-40 treatments), and most centers see only a limited number of patients. Therefore, gathering outcome data on $HBO_2$ treatment has been limited, and often, only case reports or small case series are available to support its use. In addition, practice patterns differ across centers, and some centers may use $HBO_2$ successfully for an indication that other centers may not consider it for. To gather more data on $HBO_2$ applications, outcome data from multiple centers need to be combined, but until 2011, no major academic registry existed to record treated cases and to track outcomes [5,6].

The International Registry for Hyperbaric Oxygen Treatment was formed to gather consistent outcome data from multiple centers [5]. The registry's goal is to improve the use of $HBO_2$ through evidence-based medicine. The registry records all hyperbaric medicine cases seen at participating centers and includes data on demographics, outcomes, complications, treatment duration, treatment pressure, and quality of life (QOL), among other markers. The registry started at the Dartmouth-Hitchcock Medical Center and Elliott Hospital in 2011 but expanded significantly in 2020 when the number and geographical distribution of participating centers grew. A total of 32 centers are entering data as of May 2024. The registry includes centers in the United States, the United Kingdom, and Australia (Table 1), and the data are used in publications about the use of $HBO_2$ [6-8].

An important outcome from the registry is identifying emerging medical indications for $HBO_2$ treatment. Particularly at academic centers, unique or challenging cases where hypoxia, inflammation, or both are considerations for referring patients for $HBO_2$ treatment. The types of cases referred and the outcomes from them can indicate which new indications may need further study in controlled trials. This information can also support the use of $HBO_2$ for individual conditions when other treatments are not effective. Since the registry's inception, 378 cases have been recorded where treatment was given for a condition falling outside of the current 15 UHMS-approved indications.

**Table 1.** Center totals for emerging indications contributing to the analysis (n=378).

| Center | Location | Total cases in registry (n=9019), n | Total emerging indications for $HBO_2$ therapy (n=378), n (%) |
|---|---|---|---|
| Alfred Health | Melbourne, Australia | 135 | 1 (0.3) |
| Avera McKennan Hospital | Sioux Falls, South Dakota | 310 | 10 (2.6) |
| Beverly Hospital | Beverly, Massachusetts | 67 | 2 (0.5) |
| DDRC[a] Health Care | Plymouth, United Kingdom | 109 | 4 (1.1) |
| Dartmouth-Hitchcock Medical Center | Lebanon, New Hampshire | 1118 | 74 (19.6) |
| Duke University Medical Center | Raleigh, North Carolina | 319 | 4 (1.1) |
| James Paget University Hospital | Great Yarmouth, United Kingdom | 47 | 4 (1.1) |
| Intermountain Medical Center | St Murray, Utah | 495 | 10 (2.6) |
| Latter Day Saints Hospital | Salt Lake City, Utah | 283 | 2 (0.5) |
| Legacy Health Group | Portland, Oregon | 645 | 12 (3.2) |
| LHM[b] Health Care | London, United Kingdom | 131 | 6 (1.6) |
| Logan Regional Medical Center | Logan, Utah | 110 | 1 (0.3) |
| Mayo Clinic | Rochester, Minnesota | 947 | 11 (2.9) |
| Midlands Diving Chamber | Rugby, United Kingdom | 171 | 152 (40.2) |
| Prince of Wales Hospital | Sydney, Australia | 847 | 36 (9.5) |
| Spectrum Health | Grand Rapids, Michigan | 409 | 6 (1.6) |
| St George Regional Hospital | St George, Utah | 310 | 5 (1.3) |
| St Richard's Hospital | Chichester, United Kingdom | 15 | 4 (1.1) |
| University of California San Diego | San Diego, California | 385 | 12 (3.2) |
| University of Maryland Medical Center | Baltimore, Maryland | 1260 | 29 (7.7) |
| University of Pennsylvania | Philadelphia, Pennsylvania | 223 | 1 (0.3) |
| University of Rochester Medical Center | Rochester, New York | 63 | 2 (0.5) |
| Utah Valley Hospital | Provo, Utah | 450 | 6 (1.6) |
| Wesley Hyperbaric | Brisbane, Australia | 233 | 4 (1.1) |

[a]DDRC: Diving Diseases Research Center.

[b]LHM: London Hyperbaric Medicine.

### Objectives

The goal of this analysis is to: (1) quantify which non–UHMS-approved indications are being treated at registry centers and (2) provide outcomes for indications where sufficient cases exist. This analysis is important to: (1) identify indications deserving of further research, (2) inform both hyperbaric and other practitioners on new and emerging uses of $HBO_2$, and (3) identify potential applications of $HBO_2$ treatment across various medical and surgical specialties for challenging cases.

## Methods

### Overview

To guide the presentation of data in this report, the RECORD (Reporting of Studies Conducted Using Observational Routinely Collected Health Data) statement was used [9]. RECORD was created as an extension to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement to address reporting items specific to observational studies using routinely collected health data such as registries. The checklist is included as supplemental material.

### Ethical Considerations

The structure of the web-based International Multicenter Registry for Hyperbaric Oxygen Therapy has been described previously [5]. Briefly, the registry is composed of hyperbaric centers that agree to follow the registry consortium agreement when becoming a member. Every site obtained institutional review board (IRB) approval and gathered informed consent in agreement with their IRB approval. Sites either obtain IRB approval from their own IRB or establish a reliance agreement with the IRB Dartmouth College (STUDY00024438). Informed consent differs across centers, with some centers having an approved waiver of consent and others requiring a separate consent from the hyperbaric treatment consent for all patients. Data are deidentified when they are sent to the data coordinating center. Participants do not receive compensation.

### Registry Design and Data Collection

All centers use a free database application, REDCap (Research Electronic Data Capture; Vanderbilt University) for data entry. REDCap is a secure, web-based software platform designed to support data capture for research studies, providing: (1) an intuitive interface for validated data capture, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for data integration and interoperability with external sources. Registry data from each center are anonymized and collated every quarter at the coordinating center (Dartmouth) and are available to all users for research purposes [10,11].

Data are collected within a week of beginning and ending $HBO_2$ treatment. All patients have relevant demographic data recorded (age, biological sex, and race). Yes or no questions are asked about any history of prior $HBO_2$ treatment, prior seizures, or current pregnancy. Patients are asked about diabetes (whether controlled by diet, oral medications, or insulin) and about any current or prior smoking history or other nicotine use. At the end of treatment, the number of treatments given is recorded, along with any treatment complications that may have occurred.

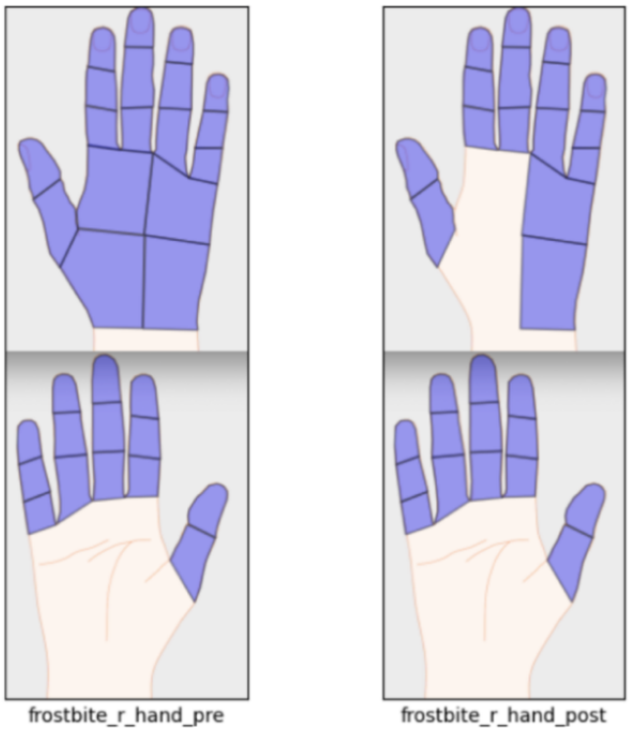
Several patient-reported outcome questionnaires are used. All the questionnaires are maintained on the registry website. Any updates to documents are thus easily distributed to the centers that all have access. Patients are asked to complete the EuroQol EQ-5D-5L QOL questionnaire at the beginning and end of treatment [12]. This questionnaire asks about health in 5 dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. Each dimension has 5 levels, and the answers to the 5 questions are combined into 1 composite score or index. Scores range from –0.59 to 1, where 1 is the best possible reported health. Overall health is also rated on a 0 to 100 hash-marked, vertical visual analog scale (VAS). The EQ-5D-5L questionnaire is a copyrighted evidence-based assessment of health and has been registered for registry use. It has been incorporated into all entries collected since 2019 to evaluate overall health pre- and post-$HBO_2$ therapy. Patients treated before then do not have results for this questionnaire.

Depending on the indication, additional questionnaires (eg, urinary distress inventory for radiation cystitis, European Organisation for Research and Treatment of Cancer, Quality of Life Questionnaire, Head and Neck 35 QLQ-H&N35 for head and neck cancer) are administered in accordance with registry protocols. For patients with Crohn disease, the perianal Crohn symptom index is completed, which asks about the number of fistulas, fistula drainage, rectal drainage, bowel movement frequency, and the use of motility agents. The administration of this questionnaire began in 2019, and it is included in Multimedia Appendix 1. Patients with ulcerative colitis (UC) receive a questionnaire that asks about bowel movement frequency during the night and day, blood in the stool, urgency, and pain. The frequency, blood, urgency, and pain factors are combined into an overall score that ranges from 0 (no symptoms) to 17 (maximum symptoms) to characterize symptom severity. The administration of this questionnaire also started in 2019, and it is included in Multimedia Appendix 2. Because the Crohn and UC questionnaires were added later in the registry's development, patients treated before that time do not have results from these questionnaires.

Patients with head trauma and post–COVID-19 symptoms are presented with the Neurobehavioral Symptom Inventory (NSI) [13]. The NSI has 22 questions designed to cover symptoms related to concussion. Patients rate the severity of common symptoms such as headache, fatigue, and nausea on a scale ranging from 0 (none) to 4 (very severe), with the maximum total score of 88. Initially, this questionnaire was only administered to patients with carbon monoxide poisoning, but in 2022, this questionnaire was expanded to any brain-related condition, including those with post–COVID-19 condition (PCC; also known as postacute sequalae of COVID-19). For frostbite, the users are presented with a diagram of the hands and feet, and they select the affected areas (Figure 1).

As the data are coded by indication, records are classified into one or more of the 15 approved UHMS indications or categorized as "other." The design of this study was to focus on all those records where the indication was "other" and to combine data for those indications. Data entries fitting into the "other" category were isolated and subsequently further divided into subgroups for investigation in this study. Data were reviewed manually to determine if the record fit into one of the already recognized "other" indications or if it was an indication that did not exist within the registry. For example, Crohn disease is a selection on the "Other Indications" dropdown menu; however, this option was not always present in the registry. Previously, users would select "other" and then enter Crohn disease in a text box. So, these text entries were examined to ensure the cases were properly classified. Once the text had been identified, data were processed using MATLAB 2023a (The MathWorks Inc) using text-recognition algorithms to classify the reasons for treatment properly. These results were manually double-checked for accuracy using Microsoft Excel. The study size was determined by the number of entries for "other" indications in the registry from registry inception in 2011 through May 2024.

**Figure 1.** Data entry screen for frostbite injuries. The registry provides different approaches to data entry to make data entry easy and consistent.




### Missing Data Reasons

Either pre- or posttreatment data were missing for a variety of reasons. For some of the questionnaires, data collection did not start until the registry had been underway for several years. This is true for the EQ-5D-5L and the Crohn, UC, and NSI questionnaires. So, patients treated early in the registry did not have results for these questionnaires. In other cases, data were missing because the patient discontinued treatment, was discharged from the hospital, or transferred to another facility and was not available for the final questionnaire. Some patients were too ill to complete the questionnaires.

### Statistical Analysis

For those indications with sufficient responses (n≥6) and complete pre- and posttreatment data for the variable in question, pre- and posttreatment values were compared statistically. Questionnaire responses or rating scales where the answers were categories were analyzed using the sign test (eg, EQ-5D-5L, NSI, and bowel questionnaire). Otherwise, data were compared using the Wilcoxon signed-rank test (eg, fistula number on the Crohn symptom index). Individuals who were missing a pre- or posttreatment value for a variable were not included in the primary analysis.

### Treatment of Missing Data

The baseline characteristics of those with complete and incomplete data on Crohn-, UC-, and PCC-related questionnaires were calculated and presented in the results. We conducted sensitivity analyses to tease out the potential impacts of missing data as follows. Individuals who were missing EQ-5D questionnaire responses but had other data in the registry indicating they had a positive response to treatment were given pre- and posttreatment values based on the pre- and posttreatment mean values of those with complete data. For the remainder, we used best, worst, and average case scenarios. In the best case, we assumed all those with missing data improved to the same degree on average as those with complete data. For the worst case, we assumed all those with missing data worsened by the same average percentage that those with complete data improved. For the average case, we calculated the proportion of individuals with complete data that improved and applied that to the missing data. We took the view that sensitivity analyses were not appropriate when the number of cases with complete data was very small, and we used 10 as the limit, accepting that this is an arbitrary limit.

## Results

### Overview

As of May 2024, a total of 32 sites were actively entering data in the registry, and 24 (75%) of these centers contributed to the indications labeled “other” (Table 1). In total, 378 cases were marked as having diagnoses not currently UHMS approved and receiving at least 1 $HBO_2$ treatment. One center had an interest in PCC and contributed 141 (94.6%) out of 149 cases for that indication. Aside from PCC, UC was the most common condition (n=47, 12.4%), followed closely by Crohn disease (n=40, 10.6%). Combined, inflammatory bowel disease (IBD; n=87) accounted for 23% of hyperbaric cases treated as “other” (n=378). If combined with pouchitis (n=1) and pyoderma gangrenosum (n=7), which are both often associated with IBD, total IBD and related cases accounted for 25.1% (95/378) of cases treated. Of the remaining diagnoses treated, calciphylaxis (n=20, 5.3%), frostbite (n=18, 4.8%), and peripheral vascular disease ulcers (n=12, 3.2%) each had ≥10 cases. There were 23 diagnoses with only a single case (Table 2). The result of this was a series of case cohorts and individual case studies gathered into a single database.

**Table 2.** Summary of other indications (n=378)[a].

| Indication | Values, n (%) |
|---|---|
| Post–COVID-19 condition | 149 (39.4) |
| Ulcerative colitis | 47 (12.4) |
| Crohn disease | 40 (10.6) |
| Calciphylaxis | 20 (5.3) |
| Frostbite | 18 (4.8) |
| Peripheral vascular disease ulcer | 12 (3.2) |
| Acute COVID-19 | 9 (2.4) |
| Pyoderma gangrenosum | 7 (1.9) |
| Pterygium | 7 (1.9) |
| Hypospadias | 7 (1.9) |
| Osteonecrosis and avascular necrosis | 6 (1.6) |
| Head trauma | 5 (1.3) |
| Infected implanted hardware | 5 (1.3) |
| Pneumatosis intestinalis | 4 (1.1) |
| Facial filler | 4 (1.1) |
| Ischemic bowel | 3 (8) |
| Raynaud syndrome | 2 (0.5) |
| Malignant otitis externa | 2 (0.5) |
| Nonarteritic anterior ischemic optic neuropathy | 2 (0.5) |
| Central retinal vein occlusion | 2 (0.5) |
| Cyclophosphamide cystitis | 2 (0.5) |
| Femoral head necrosis | 2 (0.5) |
| Invasive fungal infection | 1 (0.3) |
| Chronic anal fissure | 1 (0.3) |
| Vasculitic ulcer | 1 (0.3) |
| BK[b] virus cystitis | 1 (0.3) |
| Levamisole vasculitis | 1 (0.3) |
| Graft-vs-host disease | 1 (0.3) |
| Decubitus ulcer | 1 (0.3) |
| Greater trochanteric pain syndrome | 1 (0.3) |
| Rectovaginal fistula | 1 (0.3) |
| Argon poisoning | 1 (0.3) |
| Pouchitis | 1 (0.3) |
| Chemotherapy-related bladder ulcer | 1 (0.3) |
| After surgery in irradiated tissue | 1 (0.3) |
| Recurrent perianal abscess | 1 (0.3) |
| Tinnitus | 1 (0.3) |
| Clostridium enterocolitis | 1 (0.3) |
| Ligament and cartilage injury | 1 (0.3) |
| Branch retinal artery occlusion | 1 (0.3) |
| Axonotmesis | 1 (0.3) |

| Indication | Values, n (%) |
|---|---|
| Nonhealing bowel anastomosis | 1 (0.3) |
| Multiple sclerosis | 1 (0.3) |
| Inclusion body myositis | 1 (0.3) |
| Epidermolysis bullosa | 1 (0.3) |

[a]Some of the indications were case series with several cases; others were case reports of individual cases.

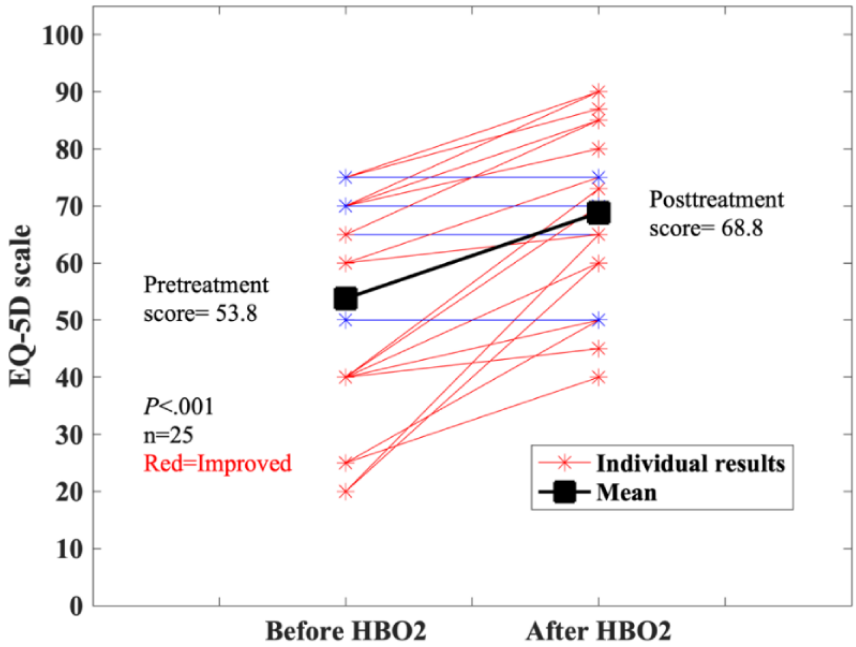
[b]BK: human polyomavirus 1.

## IBD and Related Conditions

### *Crohn Disease*

The 40 patients with Crohn disease had a mean age of 37 (SD 13) years, with 14 (35%) men and 26 (65%) women. The racial breakdown was as follows: Asian (n=2, 5%); Black (n=2, 5%), White (n=28, 70%), more than 1 race (n=1, 2%), and missing (n=7, 18%). The median number of $HBO_2$ treatments completed was 30 (IQR 19.5-40.5). A total of 25 (62%) of the 40 patients with Crohn disease had complete pre- and posttreatment data for the EQ-5D-5L QOL measure; 7 (18%) patients were missing data because they had been entered into the registry before the adoption of the questionnaire; 3 (8%) did not complete the pretreatment questionnaire, and an additional 5 (12%) did not complete the posttreatment questionnaire. Of the 3 with no pretreatment questionnaire, 1 transferred to another $HBO_2$ center and 1 was reported to have improved. Of those 5 with pretreatment data but no posttreatment questionnaire data, 1 was improving, but insurance would not approve additional treatments, 1 had closed a fistula and stopped treatment, and 1 was discharged from the hospital and could not continue as an outpatient. One completed 6 treatments and was feeling too sick to continue. Of those who completed the EQ-5D (n=25), 20 (80%) reported improvement and 5 (20%) were unchanged (*P*<.001; Figure 2). Patients with perianal Crohn questionnaire results (n=18) reported a significant decrease in discharge from the fistulas after treatment (*P*=.01; Figure 3); 5 (28%) patients also reported a decrease in the number of fistulas.

**Figure 2.** Results from the EQ-5D visual analog scale for patients with Crohn disease. A total of 20 patients had complete pretreatment and posttreatment hyperbaric oxygen (HBO2) data. A sensitivity analysis using a best, worst, and average case for missing data shows that the results are significant in all those cases.

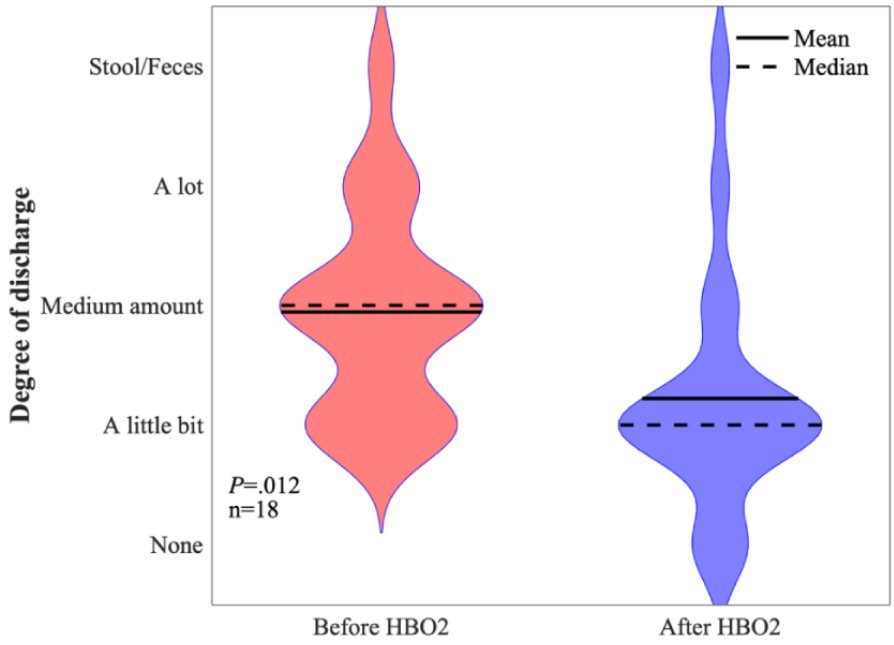



For the sensitivity analysis, the 7 individuals with missing EQ-5D questionnaires were slightly younger (32 vs 37 years) with a similar total number of treatments (n=40 vs n=37) and were considered to be missing at random. Of the remaining 8, a total of 3 (38%) were reported to have improved elsewhere in the registry and so were "given" values based on the pre- and posttreatment mean values of those with complete data, and 1 (12%) was reported to be too sick to continue, so was assumed to have worsened by the same average percentage that others had improved (28% in this case). For the remaining 4 (50%), the best, worst, and average case scenarios were used. The best case assumed that all the 4 individuals improved on the EQ-5D VAS by the average amount, and the worst case assumed that they worsened by the same average percentage that others had improved (28% in this case). For the average case, the same proportion of improvement for those with data (20/25, 80%) was applied to the 4 cases (ie, 3 improved and 1 worsened). These gave the best case (27/33, 82% improved; *P*<.001), worst case (23/33, 70% improved; *P*=.001), and average case (26/33, 79% improved; *P*<.001). If the 3 individuals with missing

EQ-5D data, but were noted to improve, were assumed to have worsened instead, this would still have been statistically significant (*P*=.04).

For the fistula discharge data (n=18), there were 18 (45%) individuals with complete data. For the sensitivity analysis, the 3 (7.5%) individuals who reported improvement elsewhere in the registry were treated the same as the others with missing data. The sensitivity analyses results are as follows: best (25/33, 76% improved; *P*<.001), worst (10/33, 30% improved; *P*=.33), and average (18/33, 55% improved; *P*=.08).

**Figure 3.** Reported fistula discharge pre- and post-HBO2 therapy in patients with Crohn disease who had complete pre- and post-HBO2 data. The sensitivity analyses results are follows: best case (25/33, 76% improved; *P*<.001), worst case 10/33, 30% improved; *P*=.33), and average case (18/33, 55% improved; *P*=.08).




### *HBO2 for UC*

The 47 patients with UC had a mean age of 41 (SD 20.5) years, with 24 (51%) women and 23 (49%) men. The racial breakdown was as follows: Asian (n=1, 2%); Black (n=2, 4%); White (n=40, 85%); refused (n=1, 2%); and missing (n=3, 6%). The median number of treatments was 5 (IQR 2.6-7.4). A total of 19 (40%) patients had complete EQ-5D-5L questionnaire data; 6 (13%) were missing data because they were in the registry before the EQ-5D was adopted; 12 (26%) had no pretreatment questionnaire data; and an additional 10 (21%) had no posttreatment questionnaire data. Of those with no pretreatment questionnaire data (n=12), 1 had cognitive impairment, 1 was unable to complete the form, 2 declined, 1 decided to stop treatment, 1 stopped because of an unrelated medical problem, and 4 stopped because they had improved. For 2 the reason was not listed. Among those with no posttreatment questionnaire data (n=10) despite having a pretreatment result, 4 (40%) were discharged before completing the form; in 3 (30%) cases, the patient terminated treatment; and in an additional 3 (30%), treatment ended unexpectedly. The group without pretreatment questionnaire data tended to be older (mean 43 years, SD 23.2) and receive more treatments (median 8.5, IQR 5.3-11.8). The group missing posttreatment questionnaire data had a similar mean age (40 years, SD 17.7) and median number of treatments (4.5, IQR 2-7) to the group having the questionnaire data. Of the 19 patients with UC who completed the EQ-5D, 3 (16%) had slightly decreased QOL scores, while most (n=16, 84.2%) had improvement (*P*=.008; Figure 4). The patients with UC who had questionnaire results also reported lower (better) scores on the bowel questionnaire (Figure 5). For the sensitivity analyses, the 4 UC cases that reported improvement were assumed to have improved by the mean of those with data. For the remainder, data are as follows: best case (38/41, 93% improved; *P*<.001), worst case (19/41, 46% improved; *P*=.76), and average case (31/41, 76% improved; *P*=.002).

There were 7 individuals with pyoderma gangrenosum, 6 (86%) women and 1 (14%) man with a mean age of 46 (SD 20.6) years. The racial breakdown was as follows: Asian (n=1, 14.3%); Black (n=1, 14.3%); and White (n=5, 71.4%). The median number of treatments was 26 (IQR 17.4-34.6). The data for these cases are limited because standard wound measures for these cases were implemented late in the registry. Only 2 (29%) cases had EQ-5D measurements; 2 (29%) cases were entered in the registry before beginning EQ-5D assessments; 2 (29%) had no pretreatment value; and 1 (14%) had no posttreatment value. A total of 3 (60%) had subjective data on improvement, and those 3 were reported to improve. One (20%) patient stopped due to an unrelated medical problem, and another (n=1, 20%) decided to stop after 7 treatments. The registry also includes 1 patient with pouchitis who improved.

**Figure 4.** Results on the EQ-5D visual analog scale for patients with ulcerative colitis (UC) who had complete pre- and posthyperbaric oxygen (HBO2) data. The sensitivity analyses results are follows: best case (38/41, 93% improved; *P*<.001), worst case (19/41, 46% improved; *P*=.76), and average case (31/41, 76% improved; *P*=.002).

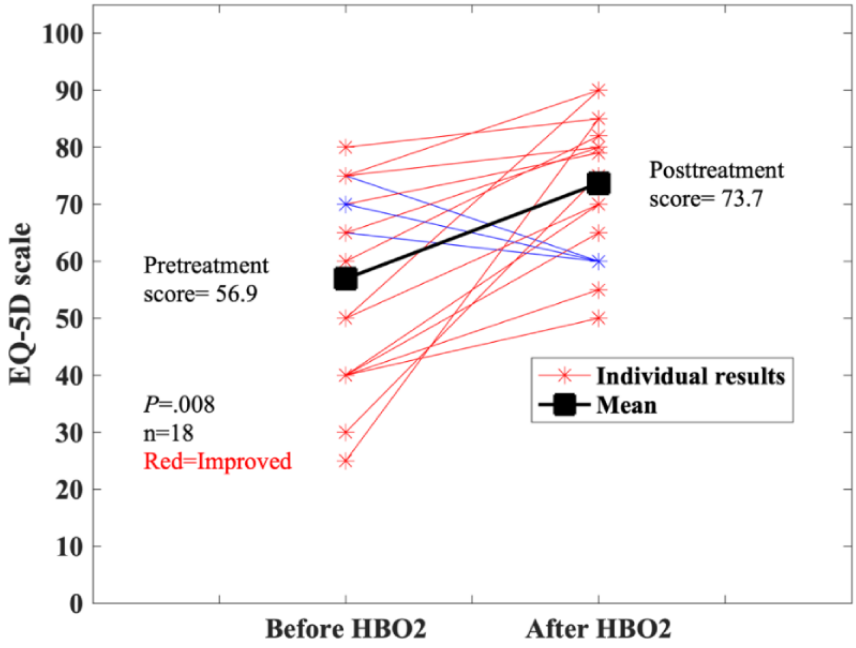
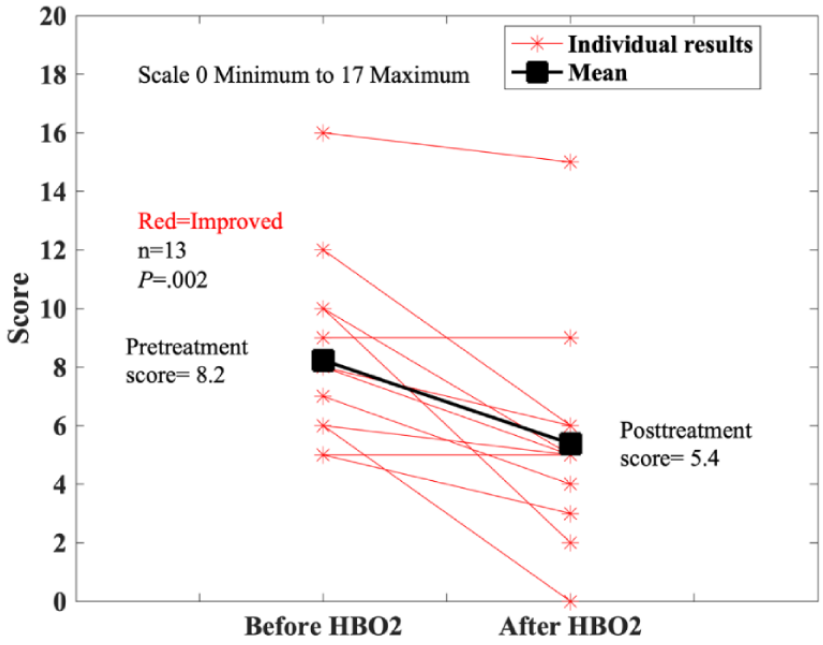



**Figure 5.** Results on the bowel questionnaire from the subset of patients with ulcerative colitis (UC) who had complete pre- and posthyperbaric oxygen (HBO2) data. The sensitivity analysis showed the best (38/41, 93%; *P*<.001) and worst (19/41, 46%; *P*=.14) cases, indicating that the results are sensitive to missing data.




### *Calciphylaxis*

Calciphylaxis was the fourth most common “other” diagnosis (n=20; Table 3). The mean age was 61 (SD 15.9), with 12 (60%) women and 8 (40%) men. The racial breakdown was as follows: Black (n=2, 10%); White (n=15, 75%); refused (n=1, 5%); and missing (n=2, 100%). The median number of treatments was 27.5 (IQR 14.5-40.5). Only 3 (15%) patients had complete QOL data, while 5 (25%) individuals did not have EQ-5D data because they were treated before the EQ-5D was in the registry, 6 (30%) had no pretreatment questionnaire, and 6 (30%) had no posttreatment questionnaire. Of the 20 patients, 4 (20%) discontinued treatment, as they were too sick to continue, and 2 of these patients died. While 2 (10%) patients healed their wounds, 5 (25%) showed notable improvement. For those who improved, the median number of treatment sessions was 39 (IQR 31.1-46.9). The median number of hyperbaric treatments in the remaining cases was 21 (IQR 7.9-34.1). Sensitivity analyses were not used because the most useful outcome for this condition is improvement versus no improvement because of the high mortality. Many of the patients missing data either died or were seriously ill.

**Table 3.** Outcomes for the patients with calciphylaxis.

| Total $HBO_2$[a] treatments | Original size (cm), L×W×D | Decrease length (%) | Decrease width (%) | Decrease depth (%) | Pretreatment EQ-5D VAS[b] | Posttreatment EQ-5D VAS | Reason for missing EQ-5D | Notes on outcome |
|---|---|---|---|---|---|---|---|---|
| **Improved** | | | | | | | | |
| 8 | 12×10×4 | 33.3 | 30 | 75 | —[c] | — | Before EQ-5D in registry | Patient decided to stop; better |
| 32 | — | — | — | — | — | — | Not done | Improved |
| 29 | 7×1×0.5 | 98.6 | 80 | 100 | — | — | Before EQ-5D in registry | Wounds healed |
| 18 | 11.3×5.4×2.5 | — | — | — | 50 | 40 | — | Wounds improving; stopped due to other health issues |
| 40 | 45×15×25 | 8.9 | 30 | 0 | 30 | — | Patient unable to return form | Despite size, considerably improved (granulation) |
| 60 | 3.7×5×0.1 | 67.6 | –22 | 10 | 87 | 92 | — | Wounds improved |
| 39 | 4.3×0.8×0.1 | 100 | 100 | 100 | — | — | Before EQ-5D in registry | Transferred to other hospital, wounds healed |
| 48 | L 1.9×5.0×2.2; R 2.0×9.5×2.2 | 58; 40 | 40; 26 | 82; 82 | — | — | Before EQ-5D in registry | Wounds mildly improved |
| 40 | — | — | — | — | 30 | 80 | — | Improvement in quality of life |
| **Not improved** | | | | | | | | |
| 17 | — | — | — | — | — | — | Unable | No change (wound care) |
| 11 | 5×2.5×1 | –500 | –500 | –100 | 10 | — | Patient did not return | Worsening or poor nutrition; stopped as futile |
| 21 | — | — | — | — | 30 | — | Unable to complete | Worsening. Too sick to continue |
| 9 | — | — | — | — | 10 | — | Stopped care | Patient decided to stop; claustrophobia noted |
| 21 | 110×100×5 | — | — | — | — | — | Unable | No change |
| 26 | 7.1×4.9×3.5 | — | — | — | 50 | — | Patient died | Transferred from outside hospital; died |
| 3 | 8×5×2 | — | — | — | — | — | Declined | Patient decided to stop (nonmedical reason) |
| 10 | — | — | — | — | — | — | Before EQ-5D in registry | Developed respiratory illness, died |
| 40 | 6×5×0 | 0 | –10 | | | | Not done | Wounds unchanged per measurements |
| 40 | — | — | — | — | — | — | Declined | No improvement |
| 40 | — | — | — | — | 40 | — | Not done | No outcomes documented |

[a]$HBO_2$: hyperbaric oxygen.

[b]VAS: visual analog scale.

[c]Not available.

### *Frostbite*

Frostbite was the fifth most common "other" diagnosis with 18 patients (Table 4). The mean age was 49 (SD 13.7) years, with 17 (94%) men and 1 (6%) woman. The racial breakdown was as follows: American Indian or Alaska Native (n=2, 11%); Black (n=1, 6%); White (n=12, 67%); other (n=2, 11%); and missing (n=1, 6%). The median number of treatments was 9 (IQR 7-11). Only 1 (6%) individual did not have EQ-5D data because the treatment occurred before the EQ-5D was in the registry, while 6 (33%) had no pretreatment questionnaire and 6 (33%) had no posttreatment questionnaire. For 2, the questionnaire was not done due to unspecified reasons; 3 either left treatment or refused treatment before completing the questionnaire; 1 had a language barrier; and 1 was not done due to inadequate staffing. For the 11 patients who had complete EQ-5D VAS data, there was a significant improvement (pretreatment score=48.2, posttreatment score=67.5; 9/11, 82% improved; $P$=.02). The sensitivity analyses results are as follows: the best case (16/18, 89% improved; $P$<.001), worst case (9/18, 50% improved; $P$=1.0), and average case (15/18, 83% improved; $P$=.002).

**Table 4.** Outcomes for patients treated for frostbite[a].

| Total $HBO_2$[b] treatments | Pretreatment EQ-5D VAS[c] | Posttreatment EQ-5D VAS | Pretreatment EQ-5D score | Posttreatment EQ-5D score | Frostbite classification | Reason for missing EQ-5D | Notes on outcome |
|---|---|---|---|---|---|---|---|
| **Improved** | | | | | | | |
| 7 | 10 | 60 | 0.16 | 0.084 | 3 | —[d] | Improved, but had amputations |
| 9 | 30 | 80 | 0.39 | 0.64 | 2 | — | No amputations, improved |
| 10 | 10 | | 0.51 | | 2 | Not done | No amputations |
| 4 | 85 | 95 | 0.33 | 0.44 | 4 | — | Initial bilateral transmetatarsal amputations, subsequently improved |
| 27 | 80 | 95 | | | | Only VAS | Had a skin graft |
| 9 | 30 | 75 | 0.69 | 0.27 | 3 | — | No amputations, good healing |
| 10 | 45 | 48 | 0.53 | 0.37 | 2 | — | Wounds stabilized |
| 7 | 50 | 50 | 0.812 | 1 | 2 | — | No amputations, wounds improved |
| 10 | 80 | 50 | 0.40 | 0.42 | 2 | — | Partial improvement, delay in $HBO_2$ due to detox |
| 7 | 50 | 70 | 0.41 | 0.80 | 2 | — | No amputations, wounds improved |
| 9 | 50 | 70 | 0.82 | 0.82 | 2 | — | Wounds improved |
| 10 | 20 | 50 | 0.06 | 0.26 | 3 | — | — |
| **Not improved or insufficient data** | | | | | | | |
| 30 | — | — | — | — | 2 | Before EQ-5D in registry | Frostbite of hands, insufficient outcome data entered |
| 15 | 0 | — | –0.074 | | 3 | Had amputations, left treatment | Amputations on right and left foot. Homeless or drug use |
| 4 | — | — | — | — | — | Not done | Bilateral below-the-knee amputations |
| 4 | — | — | — | — | 3 | Language barrier | Behavioral issues or refused treatment. Bilateral below-the-knee amputations |
| 6 | — | — | — | — | — | Staffing | Psychiatric issues or refused treatment |
| 2 | 75 | — | 0.58 | — | — | Did not return | Left treatment |

[a]The frostbite classification scale: 1=first degree; 2=second degree; 3=third degree, 4=fourth degree (includes necrosis).

[b]$HBO_2$: hyperbaric oxygen.

[c]VAS: visual analog scale.

[d]Not available.

### Peripheral Vascular Disease–Related Wounds

The sixth most commonly treated diagnosis was peripheral vascular disease–related wounds (Table 5). The mean age was 63 (SD 19.1) years, with 8 men and 4 women. The racial breakdown was as follows: 8 White, 2 Black, and 2 missing. The median number of treatments was 19.5 (IQR 8.3-30.8). Twelve cases were recorded, all involving arterial disease, but some also with concurrent venous disease. All wounds involved the lower extremities. Of those patients who had ≥20 $HBO_2$ sessions, 5 of 6 improved, with the remaining 1 lacking enough data to reach a conclusion. The remaining 6 cases had ≤15 sessions. Four of these were worse, with 2 not having sufficient data recorded. The main outcome recorded was improved versus not improved or insufficient data, and there were insufficient EQ-5D data to do a sensitivity analysis.

**Table 5.** Outcome for patients with peripheral vascular disease.

| Total $HBO_2$[a] treatments | Decrease in length (%) | Decrease in width (%) | Decrease in depth (%) | EQ-5D VAS[b] | EQ-5D VAS | EQ-5D score | EQ-5D score | Reason for missing EQ-5D | Notes on outcome |
|---|---|---|---|---|---|---|---|---|---|
| **Improved** | | | | | | | | | |
| 30 | 71.4 | 50 | 0 | __[c] | 60 | — | — | Not given pre | Improved or healing |
| 30 | 59.7 | 47.3 | 0 | 85 | 90 | 0.83 | 1.0 | — | Improved or healing |
| 30 | 66.7 | 83.3 | 83.3 | 70 | 70 | 0.68 | 0.62 | — | Improved or healing |
| 24 | 100 | 100 | 100 | — | — | — | — | Unable to collect | Resolved or healed |
| 60 | 56 | 67 | 100 | 100 | 90 | 0.42 | 0.80 | — | Stump wound closed |
| **No improvement or insufficient data** | | | | | | | | | |
| 6 | — | — | — | — | — | — | — | Unable to collect | Limited information |
| 10 | — | — | — | — | — | — | — | Before EQ-5D in registry | Worse |
| 15 | — | — | — | 75 | — | 0.88 | — | Too sick to complete | Too sick to continue |
| 1 | — | — | — | — | — | — | — | Left treatment | Patient decided to stop |
| 1 | — | — | — | — | — | — | — | Left treatment | Patient decided to stop |
| 9 | — | — | — | — | — | — | — | Left treatment | Transferred to nursing facility |
| 40 | — | — | — | — | — | — | — | Not done | Had a skin graft |

[a]$HBO_2$: hyperbaric oxygen.

[b]VAS: visual analog scale.

[c]Not available.

### Pterygium Surgery and Facial Filler Injections

One particularly interesting application of $HBO_2$ was for postoperative healing from pterygium surgery (Table 6). The mean age was 47 (SD 18.1) years, with 6 (86%) men and 1 (14%) woman, and all participants (7/7, 100%) classified as White. All cases came from a single center and were treated with 5 sessions each. One case had an unchanged QOL assessment before and after treatment but had no noted complications via notes after surgery, and 1 had worse QOL VAS scores after treatment (pretreatment score=85, posttreatment score=75), although the patient was noted to have uneventful recovery after surgery. The remaining patients (4/6, 67%) showed improvement in their QOL scores, with the overall median EQ-5D-5L VAS score increasing from 82.5 to 97.5, although this was not statistically significant ($P$=.38).

**Table 6.** Outcomes for patients with pterygium and facial filler injections.

| Total $HBO_2$[a] treatments | Pretreatment EQ-5D VAS[b] | Posttreatment EQ-5D VAS | Pretreatment EQ-5D score | Posttreatment EQ-5D score | Reason for missing EQ-5D | Notes on outcome |
|---|---|---|---|---|---|---|
| **Pterygium surgery** | | | | | | |
| 5 | 85 | 75 | 1 | 0.83 | —[c] | Uneventful |
| 5 | 90 | 90 | 0.88 | 1 | — | No postoperative problems |
| 5 | 90 | 100 | 0.86 | 1 | — | No postoperative problems |
| 5 | 80 | 100 | 0.83 | 1 | — | — |
| 5 | 70 | 100 | 0.88 | 1 | — | — |
| 5 | 70 | 95 | 0.82 | 1 | — | — |
| 5 | — | — | — | — | Not done | — |
| **Facial filler injection** | | | | | | |
| 6 | — | — | — | — | Unable | Improved |
| 1 | 95 | — | 0.81 | — | Postictal state | Had seizure |
| 5 | 75 | 85 | 0.82 | 1 | — | Improvement in bruising or ischemia |
| 5 | 95 | — | — | — | Not done | Necrosis improved with $HBO_2$ therapy |

[a]$HBO_2$: hyperbaric oxygen.

[b]VAS: visual analog scale.

[c]Not available.

$HBO_2$ was also used for complications due to the use of facial fillers (Table 6). The mean age was 46 (SD 14.5) years. All participants (4/4, 100%) were women, with 2 (50%) classified as White and 2 (50%) as missing. Of the 4 cases in the registry, 2 (50%) occurred after the use of Radiesse, a calcium hydroxylapatite filler, while the other 2 (50%) were unspecified. Of the 4 cases recorded, all were treated in a multiplace chamber, and 3 (75%) showed improvement, with 2 (50%) stopping prescribed treatment sessions early due to recovery. Both that stopped early were on a twice twice-daily schedule, with 1 at 2.0 ATA and another at 2.4 ATA. One did not have a result listed other than stopping early secondary to a seizure as a complication, but this patient was taken to 2.8 ATA, which increases seizure risk. The final case was treated at 2.4 ATA daily, completing the prescribed 5 cases with improvement. All 4 cases were treated within 1 day of referral.

## Non–Radiation-Related Osteonecrosis

Six cases of non–radiation-related osteonecrosis were treated (Table 7). The mean age was 58 (SD 17.1) years with 3 (50%) men and 3 (50%) women. The racial breakdown was as follows: White (n=4, 67%); more than 1 race (n=1, 17%); and missing (n=1, 17%). The median number of treatments was 30 (IQR 10-50). In general, these patients received long treatment courses. The improvements in QOL seen in most patients suggest that the treatment was generally successful.

**Table 7.** Outcomes for patients with non–radiation-related osteonecrosis[a].

| Total $HBO_2$[b] | Anatomic location | Pretreatment EQ-5D VAS[c] | Posttreatment EQ-5D VAS | Pretreatment EQ-5D score | Posttreatment EQ-5D score | Reason for missing EQ-5D | Notes on outcome |
|---|---|---|---|---|---|---|---|
| 20 | Knee | __[d] | — | — | — | Not asked | — |
| 60 | Ankle | 30 | 80 | 0.73 | 0.83 | — | — |
| 20 | Lower leg | 60 | 80 | 0.15 | 0.54 | — | Uneventful course |
| 60 | Left knee | 60 | 65 | — | — | Only VAS | — |
| 40 | Jaw | 35 | 50 | 0.23 | 0.57 | — | Uneventful course |
| 3 | Sinuses or palate | 50 | — | 0.76 | — | Transferred before completion | Had lymphoma on bone biopsy, $HBO_2$ suspended |

[a]Currently, no detailed outcome measures specific to osteonecrosis exist in the registry. The increase in the EQ-5D suggests improvement.

[b]$HBO_2$: hyperbaric oxygen.

[c]VAS: visual analog scale.

[d]Not available.

## Infected Implanted Hardware

Infected hardware was treated in 5 patients. The mean age was 49 (SD 9.6) years with 1 (20%) man and 4 (80%) women; 3 (60%) were White, and 2 (40%) were missing a racial classification. A total of 2 (40%) cases had knee hardware infection, and 2 (40%) cases had spinal hardware infection. All were prescribed $HBO_2$ at 2.4 ATA for 10 to 40 treatments, but none completed >30% of their course for various reasons, including patient decision, transfer to a different facility, and significant visual disturbances. None of these patients had pre- and post-QOL measurements.

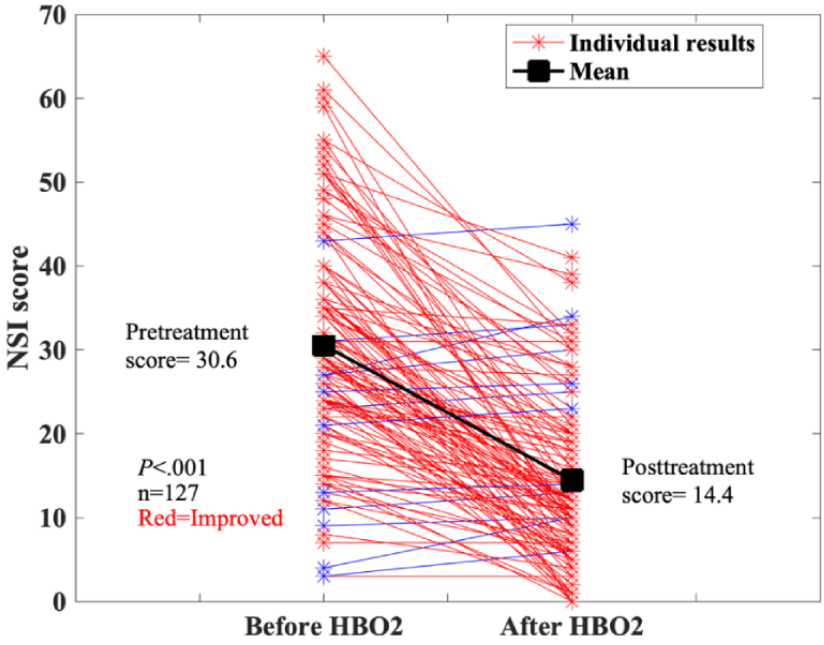
## Hypospadias

$HBO_2$ was also used in the setting of hypospadias repair. Repairs of hypospadias can lead to failed grafts with scarring, fistulas, strictures, and potential deformities [14,15]. Some centers use $HBO_2$ before and after the repair to improve graft and surgical outcomes. A total of 7 cases were entered into the registry. The mean age was 3.7 (SD 2.6) years. The racial breakdown was as follows: Asian (n=2, 29%); Black (n=1, 14%); White (n=3, 43%); and missing (n=1, 14%). For 5 (71%), the graft was preserved and viable, while for 1 (14%), the graft was preserved and partially viable and 1 (14%) was missing outcome data.

## HBO2 for PCC

A new application of $HBO_2$ was for PCC. All (n=141, 94.6%) but 8 (5.4%) of the 149 cases were reported by a single center, which has an interest in treating PCC with $HBO_2$ [16]. Complete pre- and post-$HBO_2$ NSI data were available on 127 (85.2%) patients (Figure 6). The NSI showed a significant improvement (pretreatment score=30.6, posttreatment score=14.4; $P$<.001). Complete EQ-5D data were available on 55 (36.9%) individuals. The EQ-5D VAS also improved significantly (pretreatment score=33.6, posttreatment score=64.1; $P$<.001). The group included 80 (53.7%) women and 69 (46.3%) men (total n=149); 5.4% (8/149) of patients were Asian, 0.7% (1/149) Black, 91.9% (136/149) White, 1.3% (2/149) more than 1 race and 1.3% (2/149) were missing data on race. The median number of $HBO_2$ treatments was 15 (11.5-18.5). The sensitivity analyses results are as follows: the best case (133/149, 89.3% improved; $P$<.001), worst case (111/149, 74.5% improved; $P$<.001), and average case (131/149, 87.9% improved; $P$<.001). For the EQ-5D, the cases are as follows: the best case (142/149, 95.3% improved; $P$<.001 toward improvement), worst case (49/149, 32.9% improved; $P$<.001 towards worsening), and average case (130/149, 87.2% improved; $P$<.001 toward improvement).

**Figure 6.** Results from the Neurobehavioral Symptom Inventory (NSI) for patients treated for post–COVID-19 condition. In the sensitivity analyses, the best (133/149, 89.3% improved; *P*<.001), worst (111/149, 74.5% improved; *P*<.001), and average (131/149, 87.9% improved; *P*<.001) cases all showed significant improvement.




### Other Diagnoses

For the other diagnoses, 1 group of indications was related to the relief of hypoxia (eg, ischemic bowel, Raynaud syndrome, nonarteritic anterior ischemic optic neuropathy, central retinal vein occlusion, femoral head necrosis, branch retinal artery occlusion, vasculitic ulcer). The data on chronic anal fissures and Raynaud add to the existing case series for these conditions [17,18]. Several cases focused on the treatment of cystitis (eg, human polyomavirus 1 virus cystitis, cyclophosphamide cystitis) likely motivated by the use of $HBO_2$ for radiation cystitis treatment [7]. Not enough information exists to come to conclusions about the effectiveness of the treatment for these individual cases, but this likely will change as the registry matures.

## Discussion

### Principal Findings

The registry shows how $HBO_2$ is being used across a range of medical specialties for indications where hypoxia, inflammation, or both are significant factors. The emerging diagnosis treated most across the registry centers was IBD, primarily UC and Crohn disease. For Crohn disease, there were sufficient entries to show statistically significant patient-reported improvement that remained significant even with the worst-case scenario sensitivity analysis. Overall, the newly described PCC was the most commonly treated diagnosis in this cohort, although the results were predominantly from a single center. The patient-reported results showed improvement, which supports the need for further work in this area. The top 5 indications also included calciphylaxis and frostbite, which are both conditions where hypoxia is a significant consideration and where existing treatments are often inadequate.

The results also showed indications that were concentrated at particular centers, such as the use of $HBO_2$ after complications for facial filler injections or to improve outcomes with pterygium surgery. The number of patients for each of these indications is limited but indicates that they could be studied in more detail to determine whether this approach should be adopted more widely at other centers.

### Comparison to Prior Work

#### *IBD and Related Conditions*

Several recent reviews have also concluded that $HBO_2$ could be useful for IBD [19-24]. The registry data on Crohn disease support other studies showing improvement with $HBO_2$ [19,25-29]. For fistulizing Crohn disease, a prospective trial has shown reduced disease activity, reduced drainage, and clinical remission when $HBO_2$ is used in refractory cases [26]. The improvement in the registry of Crohn cases is noteworthy because Crohn disease is not currently a UHMS-approved indication for $HBO_2$, and patients are almost always referred because standard treatments have not been successful. The patients reported significant improvement in QOL and reduced fistula and rectal discharge. Five patients had their fistulas closed. The QOL results were significant even with the worst-case approach to the sensitivity analysis. The fact that 5 patients were able to close fistulas is significant because these were likely refractory patients. Taken together, the results add to a growing set of case series and case studies that show improvement in difficult Crohn cases and support the need for large-scale trials.

For UC, $HBO_2$ therapy studies have shown clinical remission, avoidance of colectomy, or progression to second-line therapy in a randomized sham-controlled trial [30]. A follow-on phase 2B trial examined the dosing strategies and concluded that 5 sessions were superior to 3 for a UC flare [31]. Typically, $HBO_2$

is used adjunctively in UC with other treatments (steroids, biological agents, etc), although in 1 case study, $HBO_2$ was used successfully as monotherapy for UC in an individual who had contraindications to most standard treatments [32]. $HBO_2$ can be particularly useful in situations where the individual cannot receive standard treatment due to allergy, antibodies, shingles, or other reasons. The most common application of $HBO_2$ in IBD is patients admitted for an acute UC flare who are at risk for colectomy or who cannot receive steroids or biological agents or where there is a delay in being able to start those agents [30,31]. In addition, patients admitted for an acute UC flare who need a bridge between when IV steroids are begun and when biological agents (ie, infliximab) start to have an effect may benefit from $HBO_2$. These registry results support the idea that $HBO_2$ can be useful in UC but should be interpreted with caution. Many of the patients were hospitalized for treatment and were receiving intravenous steroids and other interventions. So, while the reports of improvement are encouraging, separating the effect of $HBO_2$ from the effects of standard treatment is not possible in this data set. In addition, the fact that 1 center (Dartmouth-Hitchcock) accounted for many of the treated cases (roughly 75%) and had been involved in an UC clinical trial may overrepresent the interest in UC cases across the registry centers. Also, the QOL and bowel questionnaire results were not significant in the worst-case scenario of the sensitivity analysis. Randomized controlled trials are needed to determine definitively if $HBO_2$ improves on standard care.

In addition to UC and Crohn disease, IBD-related conditions where hyperbaric treatment could be applied include pouchitis, nonhealing ileal pouch-anal anastomoses, and pyoderma gangrenosum [33-37]. Two published case series and a case report show relief of symptoms and decreased disease activity in pouchitis [33,34,38]. Case series and case studies also show reduced wound size in pyoderma gangrenosum [39]. The results from this study align with these conclusions, but the registry results are descriptive and do not include a control group for comparison (as is true for most case reports and case series).

The IBD results support the idea that $HBO_2$ could be considered in UC flares when standard therapy, such as steroids and biologics, is inadequate or not possible. Similarly, it could be useful in fistulizing perineal Crohn disease refractory to medical and surgical treatments. Pouchitis that has not responded to antibiotics and immunomodulators as standard therapy and pyoderma gangrenosum refractory to steroids and immunologics are also potential applications of $HBO_2$. IBD may be a disease process that should be evaluated as a UHMS-approved indication for $HBO_2$, as accumulating evidence shows promising results.

### *Calciphylaxis*

Calciphylaxis was the fourth most treated diagnosis. Calcification of small blood vessels leads to hypoxic painful wounds, which led to the use of $HBO_2$ in these patients. No randomized control trials exist to guide the use of possible calciphylaxis interventions. Such trials would likely be very difficult to carry out because the disease is rare and disease severity varies widely. Also, calciphylaxis has a very high mortality. Most patients with calciphylaxis have end-stage renal disease (ESRD), and the 1-year mortality for those with ESRD and calciphylaxis is 45% to 80%; 1-year mortality is less but still significant for those without ESRD (25%-45%) [40]. On the basis of studies done to date, $HBO_2$ has been recommended by subject matter experts to facilitate healing of recalcitrant calciphylactic ulcers after accounting for cost, availability, and patient tolerance of treatment [41]. Physiologically, $HBO_2$ addresses the tissue hypoxia caused by the damage to small blood vessels, which is characteristic of calciphylaxis. $HBO_2$ greatly increases the amount of oxygen in circulation reaching hypoxic wounds, which promotes collagen formation and stimulates angiogenesis.

The results from this study support what has been seen in prior work. A retrospective study of $HBO_2$ therapy for calciphylaxis showed promising results; 34 patients received a full course of $HBO_2$, with 58% (n=20) of patients improving and 32% (n=11) completely healing their wounds [42]. In this prospective registry study, 45% (9/20) of patients improved. Similar to other studies, the registry results for calciphylaxis are mixed, which is not surprising considering the nature of this disease. Affected patients often have significant medical comorbidities, and the disease has a high morbidity and mortality rate secondary to sepsis. For example, in this cohort, 2 patients treated died of their disease, and 2 others were too sick to continue.

Nevertheless, there were several cases that improved significantly, and these successful cases had more sessions of therapy. Given the severe mortality and morbidity associated with the disease and the difficult course of standard care, $HBO_2$ is a treatment that has little risk with the potential for significant benefit. The results support previous positive case series and suggest that trials on $HBO_2$ and calciphylaxis are needed, and more research should be made to pursue calciphylaxis as a possible UHMS indication.

### *Frostbite*

Frostbite causes hypoxia and necrosis in affected tissues, and $HBO_2$ has been considered as a potential therapy for many years. Kemper et al [43] summarized the literature on frostbite and $HBO_2$ in 2014. In 2021, Magnan et al [44] published a multicenter prospective single-arm study of individuals with stage 3 or 4 frostbite who received $HBO_2$ in addition to iloprost (n=28). The results were compared with a historical cohort that received iloprost alone (n=30). The results showed a significantly higher proportion of preserved frostbitten segments in the $HBO_2$+iloprost group. The registry results support the results from these previous studies. Most individuals with EQ-5D data reported an improved QOL. In the sensitivity analysis, this was significant in the best and average cases but not the worst case. Determining whether $HBO_2$ offers an additional benefit over standard treatment is not possible within the current registry design, but future expansion efforts could include trying to include registry centers without $HBO_2$ that treat similar cases.

### *Peripheral Vascular Disease–Related Wounds*

Many of the emerging indications shared the underlying problem of tissue hypoxia. For the same reason that $HBO_2$ therapy would work in treating acute arterial insufficiency, it might assist in the healing of chronic peripheral arterial disease ulcers. A guideline from a wound care advisory panel for the treatment of arterial insufficiency ulcers recommends adjuvant $HBO_2$ for patients with nonreconstructable anatomy or whose ulcer is not healing despite revascularization. The guideline also recommends determining if the hypoxia is reversible by $HBO_2$ therapy [45].

The registry cases show that some people did respond. Successful cases had at least 30 treatment sessions, while those with <15 sessions generally did not improve. Detailed information regarding each case is not available within the registry, so it is not known whether those who improved did so because they received more treatments or if they received more treatments because they had fewer comorbidities or were showing early benefits. In addition, whether the patients were screened using in-chamber transcutaneous oxygen measurements to document a local tissue increase in oxygen levels with hyperbaric oxygenation is not recorded. Patient selection is probably critical for deriving the most benefit from $HBO_2$ in these patients because those with very severe hypoxia and vascular compromise may not respond. Further studies are needed to define how to select patients with peripheral vascular disease that might benefit from $HBO_2$. In addition, information on screening procedures used to select these patients for treatment should be added to the registry.

### *Other Applications*

An interesting application of $HBO_2$ that was concentrated at 1 center, was the use of $HBO_2$ after pterygium surgery. The recurrence rate of simple surgical excision can be as high as 40% but can be lowered to 16% with a conjunctival autograft procedure, although the recurrence rates are increased if the pterygium is already recurrent when the procedure is used [46]. The use of postoperative $HBO_2$ treatment is based on a prospective trial published in 2011, which showed low recurrence rates when patients with pterygium underwent surgery with adjunctive $HBO_2$ afterward [47]. A short 5-session $HBO_2$ prescription was used. There were no complications noted with $HBO_2$ in these cases and, importantly, no reported recurrences of pterygium. This local protocol may be feasible to attempt at other centers, should collaboration with ophthalmology be obtained. At the very least, this is an interesting condition to further evaluate for using $HBO_2$ as a successful adjunct treatment to standard therapy. Another indication that is also focused on improving outcomes after surgery is the use of $HBO_2$ for individuals who have had hypospadias repairs. There were 7 individuals in the registry with this indication, and all reported preservation of the graft, although in 1, the graft was partially viable. These results are consistent with prior work [14]. In addition, this application could be considered as an extension of the approved UHMS indication of compromised flaps and grafts.

Facial filler complications present an emergency treatment opportunity well suited for $HBO_2$ and have been reported in several case reports [48-51]. This was identified in the registry cases as well. The experience of providers performing these outpatient facial filler procedures varies widely as does the ability and knowledge to address the complications. With rising interest in esthetic therapy, complications from these procedures are becoming more common. Facial fillers typically refer to hyaluronic acid fillers or particulate-based fillers, both of which are used for facial remodeling or tissue augmentation for cosmetic or reconstructive purposes. These injections can cause vascular obstruction and subsequent tissue ischemia; thus, the goal of treatment is to restore perfusion to the affected area. Although hyaluronic acid can be reversed with hyaluronidase, particulate injections such as calcium hydroxylapatite cannot easily be broken down, and $HBO_2$ may play an even greater role in these cases. Expert consensus protocols include $HBO_2$ in treatment algorithms for these complications [52]. Similar to acute arterial ischemic cases, success likely hinges on the hyperbaric medicine consult occurring close to the initial injury. Early intervention must be encouraged, as with other UHMS-approved indications where acute ischemia is prominent such as central retinal artery occlusion or compromised flaps, because earlier intervention leads to improved results. Half of the cases recorded were treated on the same day as the consult, although information on time elapsed since filler injection is not available. Ultimately, $HBO_2$ can support threatened tissue but cannot resuscitate dead tissue, and aggressive intervention and resuscitation should be advocated. Information about the use of $HBO_2$ for this indication needs to be available to other centers that administer facial filler injections.

When implanted hardware is infected, the medical standard of care typically calls for hardware removal. Removal from certain areas of the body (eg, brain and spine) can be difficult or involve complicated surgical intervention. In these instances, salvage of hardware has the lowest risk-to-benefit ratio or may be the only viable option [53]. Ridding foreign bodies of infection in vivo is notoriously difficult, which is why $HBO_2$ was likely attempted in these cases. None of the patients treated completed their course of treatment, which may suggest that these cases were refractory. The results suggest that the registry should be revised to collect more detailed information about these cases.

The newly recognized PCC was treated multiple times in the registry. The anti-inflammatory effects of $HBO_2$ are thought to be the mechanism of action underlying improvement. PCC is not yet completely understood but likely results from a dysregulated inflammatory process within the brain [54]. In general, patients reported improvement with treatment. This matches reports coming from case reports and series, as well as from a randomized control trial out of Israel [55]. The results need to be interpreted with caution, as placebo effects with $HBO_2$ are common. Nevertheless, the data support further sham-controlled trials and suggest that $HBO_2$ could be used to treat other postinfectious syndromes in the future.

Many of the other individual case reports within the registry are for indications where hypoxia is believed to be a significant

factor. For example, femoral head necrosis results from a compromised vascular supply, and healing of chronic anal fissures is believed to be hampered by local hypoxia [56]. Ophthalmological conditions such as branch retinal artery occlusion, retinal vein occlusion, and nonarteritic anterior ischemic neuropathy involve hypoxia and a compromised blood supply. Raynaud syndrome and vasculitic ulcers are also characterized by local hypoxia. Currently, not enough data exist within the registry to draw conclusions about outcomes for these various conditions, but over time, additional cases will accumulate.

The hyperbaric registry is still in its nascent stages. Greater adoption of the registry by the community will lead to more significant results. The cases highlighted in the report have identified additional data that should be recorded for some of these indications. For example, facial filler cases highlight the need for further information such as the time of injury to help determine the optimal timing for maximal $HBO_2$ efficacy. In addition, the facial filler cases should be documented in an identified category, as is done with indications such as UC or Crohn disease, rather than just marked as "other." More specific outcome measures are needed for non-radiation–related osteonecrosis cases. In addition, long-term follow-up will be essential for infected hardware cases because the outcome may not be known until well after the treatments have stopped.

Many of the involved centers are academic, nonprofit, and governmental entities with an interest in advancing evidence-based medicine and research. The contributions made by such centers are essential to evaluating the promise of $HBO_2$ in various conditions and may lead to the adoption of them as new UHMS indications. This review provides a snapshot to clinicians of the diagnoses that have been treated by $HBO_2$ that are not officially on the UHMS indications list. This can be useful in situations where patients are not improving with standard therapy, and other approaches are being sought that could address hypoxia and inflammation. It also focuses attention on specific diagnoses worthy of further study.

## Limitations and Strengths of This Study

The registry has limitations. To keep the time required for data entry to a minimum, only a specific set of outcomes is collected. The registry does not include extensive data on medications, other comorbid conditions, or laboratory results. While this approach has been successful at making the registry practical for centers to adopt, it does affect the scope of conclusions that can be drawn.

The registry undergoes active review and expansion over time, so not all questionnaires currently in use existed in the registry when it started. Missing data, particularly for the EQ-5D in earlier cases, is understandable, as this questionnaire was added later in registry development. In addition, data collection can be missed if patients stop treatment early or unexpectedly or transfer to another center. Nevertheless, the missing data are a limitation of the registry. The registry does not include a control or standard treatment group for every condition, so it is not possible to compare outcomes with no treatment or placebo. Registry results can be compared with publicly reported treatment success rates of those using standard therapy.

For the analyses, the data were reviewed to fix indications that were misclassified (ie, UHMS-approved indications that were categorized as "other" or "other" indications that were inappropriately classified as UHMS-approved indications). There may be cases, however, that were missed. The data are dependent on data entry, and there may be variations in data entry quality between centers. Work is ongoing with data quality checks throughout the registry centers to ensure that these problems are limited in the future.

Nevertheless, the strength of the registry is that it allows outcomes from less common conditions to be combined across centers, which provides information that no individual center would be able to achieve independently. The accumulated results identify areas where more detailed trials would be useful. In addition, some centers are beginning long-term follow-up with treated patients to gather data on satisfaction with and longevity of the results. This will expand the capabilities of the registry although the responses will be affected by selection bias.

## Conclusions

The registry is providing standardized outcome data on conditions that are not currently UHMS approved but are being treated with $HBO_2$ at participating centers. This is adding prospectively collected data to the mostly retrospective data that exist in the literature. Results show statistically significant improvements in patient-reported outcomes, such as QOL, for IBD, frostbite, and PCC. $HBO_2$ is also being used to improve surgical outcomes for pterygium and hypospadias procedures and to treat complications from facial filler injections. $HBO_2$ is also being tried for difficult cases involving hypoxia such as femoral head necrosis, Raynaud syndrome, and chronic anal fissures.

As time goes on and as more centers participate, significant trends can be identified, which may identify a valuable treatment option for medical conditions where hypoxia and inflammation are important contributing factors. Indications developed at a few centers initially can also be expanded to other centers based on the outcomes.

## Acknowledgments

The authors at the individual registry sites of the Multicenter Registry for Hyperbaric Oxygen Treatment Consortium are: Bridget Devaney, MBChB, FACEM Alfred Health, Melbourne, Australia; Brian Pruss, RRT/CHT, Avera McKennan Hospital, Sioux Falls, South Dakota; George Kacoyanis, MD, Beverly Hospital, Beverly, Massachusetts; Mark Bettley-Smith, MBBS, MRCGP, DRCOG, The Diver Clinic, Poole, the United Kingdom; Gary Smerdon, PhD, DDRC Healthcare, Hyperbaric Medical Centre, Plymouth, the United Kingdom; St George Regional Medical Center, St. George, Utah; Richard Moon, MD, Duke University

Medical Center, Durham, North Carolina; Pieter Bothma, MD, East of England Hyperbaric Unit, James Paget University Hospital, Great Yarmouth, the United Kingdom, and LHM Healthcare Hyperbaric Unit, Whipps Cross University Hospital, London, the United Kingdom (LHM); Renee Woodring, RN, Elliot Health System, Manchester, New Hampshire; Hyperbaric Medicine Unit, St Richard's Hospital, Chichester, the United Kingdom; Lindell Weaver, MD, Intermountain Medical Center, Salt Lake City, Utah; Latter Day Saints Hospital, Salt Lake City, Utah; Enoch Huang, MD, Legacy Medical Group, Portland, Oregon; Logan Regional Hospital, Logan, Utah; Gary Toups, MD, Mayo Clinic, Rochester, Minnesota; McKay-Dee Hospital, Ogden, Utah; Midlands Diving Chamber, Rugby, the United Kingdom; Bruce Mathew, MD and Gerard Laden, North England Medical and Hyperbaric Services, Hull, the United Kingdom; Northwest Recompression Unit, Birkenhead, the United Kingdom; Michael H Bennett, MBBS, FANZCA, MD and Glen C Hawkins, MBChB, FANZCA, Prince of Wales Hospital, Sydney, Australia, St Alphonsus Hospital System, Boise, Idaho; St. Luke's Health System, Boise, Idaho; Ian Grover, MD and Hideaki L Tanaka, MD, University of California at San Diego, San Diego, California; Melissa Schroder and Kinjal Sethuraman, MD, University of Maryland Medical Center, Baltimore; Zachary Gaskill, DO. University of Pennsylvania Health System, Philadelphia, Pennsylvania; Tammy Roman, EdD, MSN, RN, NEA-BC and Jennifer Wright, RN, BSN, CHRN, University of Rochester Medical Center, Rochester, New York; Marc Robins, DO, Utah Valley Hospital, Provo, Utah (UVH); Susannah Sherlock, MBBS, BSc, FANZCA, Wesley Hyperbaric, Auchenflower, AU; Bolaji Coker manages the United Kingdom central registry at King's College London.

The registry receives support from the Undersea and Hyperbaric Medical Society and from the Clement F Burnap Endowment to Support the Pressure Chambers for Clinical Medicine and Research at Dartmouth Medical School. The Department of Medicine at the Dartmouth-Hitchcock Medical Center provided funding support for the registry through the Department of Medicine Scholarship Enhancement in Academic Medicine Award Program. The Washington University School of Medicine also provided funding support. The REDCap instance at Dartmouth receives support from the Biomedical Informatics Core grant (principal investigator Sergio Duncan).

## Data Availability

The data and code used for the study are available from the Steering Committee of the Multicenter Registry for Hyperbaric Oxygen Treatment upon reasonable request.

## Authors' Contributions

HLT contributed to investigation, formal analysis, original draft preparation, and manuscript review and editing; JRR contributed to conceptualization, methodology, supervision, and manuscript review and editing; ZZ contributed to data curation and formal analysis; JAP contributed to conceptualization, investigation, and statistical support; PMH contributed to investigation and methodology; EMS contributed to investigation and manuscript review and editing; JLP contributed to methodology, supervision, formal analysis, and manuscript review and editing; and JCB contributed to conceptualization, funding acquisition, formal analysis, supervision, original draft preparation, manuscript review, and editing.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Perianal Crohn Symptom Index.
[PDF File (Adobe PDF File), 144 KB - ijmr_v13i1e53821_app1.pdf ]

Multimedia Appendix 2
Bowel Symptoms Questionnaire.
[PDF File (Adobe PDF File), 81 KB - ijmr_v13i1e53821_app2.pdf ]

## References

1. Thom SR. Hyperbaric oxygen: its mechanisms and efficacy. Plast Reconstr Surg 2011 Jan;127 Suppl 1(Suppl 1):131S-141S [FREE Full text] [doi: 10.1097/PRS.0b013e3181fbe2bf] [Medline: 21200283]
2. Thom SR, Bhopale VM, Velazquez OC, Goldstein LJ, Thom LH, Buerk DG. Stem cell mobilization by hyperbaric oxygen. Am J Physiol Heart Circ Physiol 2006 Apr;290(4):H1378-H1386 [FREE Full text] [doi: 10.1152/ajpheart.00888.2005] [Medline: 16299259]
3. Weaver LK. Hyperbaric oxygen therapy indications. Thirteenth edition. Undersea and Hyperbaric Medical Society. 2014. URL: https://www.uhms.org/images/indications/UHMS_HBO2_Indications_13th_Ed._Front_Matter__References.pdf [accessed 2024-08-05]
4. Butler WP. Caisson disease during the construction of the Eads and Brooklyn Bridges: a review. Undersea Hyperb Med 2004;31(4):445-459. [Medline: 15686275]

5. Harlan NP, Ptak JA, Rees JR, Cowan DR, Fellows AM, Kertis JA, et al. Development of an international, multicenter, hyperbaric oxygen treatment registry and research consortium: protocol for outcome data collection and analysis. JMIR Res Protoc 2020 Aug 17;9(8):e18857 [FREE Full text] [doi: 10.2196/18857] [Medline: 32579537]
6. Harlan NP, Ptak JA, Rees JR, Cowan DR, Fellows AM, Moses RA, et al. International multicenter registry for hyperbaric oxygen therapy: results through June 2021. Undersea Hyperb Med 2022;49(3):275-287. [Medline: 36001560]
7. Moses RA, Hunter AE, Brandes ER, Zhang Z, Rees JR, Peacock JL, et al. Patient-reported outcome measures following hyperbaric oxygen therapy for radiation cystitis: early results from the multicenter registry for hyperbaric oxygen therapy. J Urol 2024 Jun;211(6):765-774. [doi: 10.1097/JU.0000000000003929] [Medline: 38573938]
8. Sedlacek M, Harlan NP, Buckey JCJ. Renal effects of hyperbaric oxygen therapy in patients with diabetes mellitus: a retrospective study. Int J Nephrol 2021 Jun 12;2021:9992352 [FREE Full text] [doi: 10.1155/2021/9992352] [Medline: 34234965]
9. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies conducted using observational routinely-collected health data (RECORD) statement. PLoS Med 2015 Oct 06;12(10):e1001885 [FREE Full text] [doi: 10.1371/journal.pmed.1001885] [Medline: 26440803]
10. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
11. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
12. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
13. Soble JR, Silva MA, Vanderploeg RD, Curtiss G, Belanger HG, Donnell AJ, et al. Normative Data for the Neurobehavioral Symptom Inventory (NSI) and post-concussion symptom profiles among TBI, PTSD, and nonclinical samples. Clin Neuropsychol 2014;28(4):614-632. [doi: 10.1080/13854046.2014.894576] [Medline: 24625213]
14. Chua ME, Kim JJ, Ming JM, De Jesus MJ, See MC4, Bagli DJ, et al. The utilization of hyperbaric oxygenation therapy in hypospadias repair: a systematic review and meta-analysis. Int Urol Nephrol 2022 Feb;54(2):273-285. [doi: 10.1007/s11255-021-03096-y] [Medline: 35024997]
15. Neheman A, Rappaport YH, Verhovsky G, Bush N, Snodgrass W, Lang E, et al. Hyperbaric oxygen therapy for pediatric "hypospadias cripple" - evaluating the advantages regarding graft-take. J Pediatr Urol 2020 Aug;16(4):517-518. [doi: 10.1016/j.jpurol.2020.05.149] [Medline: 32576432]
16. Robbins T, Gonevski M, Clark C, Baitule S, Sharma K, Magar A, et al. Hyperbaric oxygen therapy for the treatment of long COVID: early evaluation of a highly promising intervention. Clin Med (Lond) 2021 Nov;21(6):e629-e632 [FREE Full text] [doi: 10.7861/clinmed.2021-0462] [Medline: 34862223]
17. Cundall JD, Gardiner A, Laden G, Grout P, Duthie GS. Use of hyperbaric oxygen to treat chronic anal fissure. Br J Surg 2003 Apr;90(4):452-453. [doi: 10.1002/bjs.4054] [Medline: 12673747]
18. Dowling GB, Copeman PW, Ashfield R. Raynaud's phenomenon in scleroderma treated with hyperbaric oxygen. Proc R Soc Med 1967 Dec;60(12):1268-1269 [FREE Full text] [Medline: 6066573]
19. Feitosa MR, Féres Filho O, Tamaki CM, Perazzoli C, Bernardes MV, Parra RS, et al. Adjunctive hyperbaric oxygen therapy promotes successful healing in patients with refractory Crohn's disease. Acta Cir Bras 2016;31 Suppl 1:19-23 [FREE Full text] [doi: 10.1590/S0102-86502016001300005] [Medline: 27142900]
20. You JH, Jiang JL, He WB, Ma H, Zhou M, Chen XX, et al. Addition of hyperbaric oxygen therapy versus usual care alone for inflammatory bowel disease: a systematic review and meta-analysis. Heliyon 2022 Oct 13;8(10):e11007 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e11007] [Medline: 36276722]
21. Chen P, Li Y, Zhang X, Zhang Y. Systematic review with meta-analysis: effectiveness of hyperbaric oxygenation therapy for ulcerative colitis. Therap Adv Gastroenterol 2021 Jul 17;14:17562848211023394 [FREE Full text] [doi: 10.1177/17562848211023394] [Medline: 34349835]
22. McCurdy J, Siw KC, Kandel R, Larrigan S, Rosenfeld G, Boet S. The effectiveness and safety of hyperbaric oxygen therapy in various phenotypes of inflammatory bowel disease: systematic review with meta-analysis. Inflamm Bowel Dis 2022 Mar 30;28(4):611-621. [doi: 10.1093/ibd/izab098] [Medline: 34003289]
23. Singh AK, Jha DK, Jena A, Kumar- M, Sebastian S, Sharma V. Hyperbaric oxygen therapy in inflammatory bowel disease: a systematic review and meta-analysis. Eur J Gastroenterol Hepatol 2021 Dec 01;33(1S Suppl 1):e564-e573. [doi: 10.1097/MEG.0000000000002164] [Medline: 33905214]
24. Wu X, Liang TY, Wang Z, Chen G. The role of hyperbaric oxygen therapy in inflammatory bowel disease: a narrative review. Med Gas Res 2021;11(2):66-71 [FREE Full text] [doi: 10.4103/2045-9912.311497] [Medline: 33818446]
25. Lansdorp CA, Buskens CJ, Gecse KB, D'Haens GR, Van Hulst RA. Wound healing of metastatic perineal Crohn's disease using hyperbaric oxygen therapy: a case series. United European Gastroenterol J 2020 Aug;8(7):820-827 [FREE Full text] [doi: 10.1177/2050640620934915] [Medline: 32529922]

26. Lansdorp CA, Gecse KB, Buskens CJ, Löwenberg M, Stoker J, Bemelman WA, et al. Hyperbaric oxygen therapy for the treatment of perianal fistulas in 20 patients with Crohn's disease. Aliment Pharmacol Ther 2021 Mar;53(5):587-597 [FREE Full text] [doi: 10.1111/apt.16228] [Medline: 33326623]
27. Dulai PS, Gleeson MW, Taylor D, Holubar SD, Buckey JC, Siegel CA. Systematic review: the safety and efficacy of hyperbaric oxygen therapy for inflammatory bowel disease. Aliment Pharmacol Ther 2014 Jun;39(11):1266-1275 [FREE Full text] [doi: 10.1111/apt.12753] [Medline: 24738651]
28. Colombel JF, Mathieu D, Bouault JM, Lesage X, Zavadil P, Quandalle P, et al. Hyperbaric oxygenation in severe perineal Crohn's disease. Dis Colon Rectum 1995 Jun;38(6):609-614. [doi: 10.1007/BF02054120] [Medline: 7774472]
29. Noyer CM, Brandt LJ. Hyperbaric oxygen therapy for perineal Crohn's disease. Am J Gastroenterol 1999 Feb;94(2):318-321. [doi: 10.1111/j.1572-0241.1999.00848.x] [Medline: 10022622]
30. Dulai PS, Buckey JCJ, Raffals LE, Swoger JM, Claus PL, O Toole K, et al. Hyperbaric oxygen therapy is well tolerated and effective for ulcerative colitis patients hospitalized for moderate-severe flares: a phase 2A pilot multi-center, randomized, double-blind, sham-controlled trial. Am J Gastroenterol 2018 Oct;113(10):1516-1523. [doi: 10.1038/s41395-018-0005-z] [Medline: 29453383]
31. Dulai PS, Raffals LE, Hudesman D, Chiorean M, Cross R, Ahmed T, et al. A phase 2B randomised trial of hyperbaric oxygen therapy for ulcerative colitis patients hospitalised for moderate to severe flares. Aliment Pharmacol Ther 2020 Sep;52(6):955-963. [doi: 10.1111/apt.15984] [Medline: 32745306]
32. Harlan NP, Roberts J, Siegel C, Buckey JC. Hyperbaric oxygen as successful monotherapy for a severe ulcerative colitis flare. Inflamm Bowel Dis 2022 Sep 01;28(9):1474-1475. [doi: 10.1093/ibd/izac141] [Medline: 35771656]
33. Hasan B, Yim Y, Ur Rashid M, Khalid RA, Sarvepalli D, Castaneda D, et al. Hyperbaric oxygen therapy in chronic inflammatory conditions of the pouch. Inflamm Bowel Dis 2021 Jun 15;27(7):965-970. [doi: 10.1093/ibd/izaa245] [Medline: 32944766]
34. Fahad H, Dulai PS, Shen B, Kochhar GS. Hyperbaric oxygen therapy is effective in the treatment of inflammatory and fistulizing pouch complications. Clin Gastroenterol Hepatol 2021 Jun;19(6):1288-1291. [doi: 10.1016/j.cgh.2020.06.029] [Medline: 32565291]
35. Seo HI, Lee HJ, Han KH. Hyperbaric oxygen therapy for pyoderma gangrenosum associated with ulcerative colitis. Intest Res 2018 Jan;16(1):155-157 [FREE Full text] [doi: 10.5217/ir.2018.16.1.155] [Medline: 29422812]
36. Zeina T, Levy AN. Adjunctive hyperbaric oxygen therapy for high risk peristomal pyoderma gangrenosum. Dig Dis Sci 2022 Oct;67(10):4597-4598. [doi: 10.1007/s10620-022-07646-x] [Medline: 35989384]
37. Hill DS, O'Neill JK, Toms A, Watts AM. Pyoderma gangrenosum: a report of a rare complication after knee arthroplasty requiring muscle flap cover supplemented by negative pressure therapy and hyperbaric oxygen. J Plast Reconstr Aesthet Surg 2011 Nov;64(11):1528-1532. [doi: 10.1016/j.bjps.2011.03.024] [Medline: 21514259]
38. Nyabanga CT, Kulkarni G, Shen B. Hyperbaric oxygen therapy for chronic antibiotic-refractory ischemic pouchitis. Gastroenterol Rep (Oxf) 2017 Nov;5(4):320-321 [FREE Full text] [doi: 10.1093/gastro/gov038] [Medline: 26319238]
39. Tutrone WD, Green K, Weinberg JM, Caglar S, Clarke D. Pyoderma gangrenosum: dermatologic application of hyperbaric oxygen therapy. J Drugs Dermatol 2007 Dec;6(12):1214-1219. [Medline: 18189061]
40. Nigwekar SU, Thadhani R, Brandenburg VM. Calciphylaxis. N Engl J Med 2018 Jul 26;379(4):399-400. [doi: 10.1056/NEJMc1807324] [Medline: 30044942]
41. Nigwekar SU, Kroshinsky D, Nazarian RM, Goverman J, Malhotra R, Jackson VA, et al. Calciphylaxis: risk factors, diagnosis, and treatment. Am J Kidney Dis 2015 Jul;66(1):133-146 [FREE Full text] [doi: 10.1053/j.ajkd.2015.01.034] [Medline: 25960299]
42. An J, Devaney B, Ooi KY, Ford S, Frawley G, Menahem S. Hyperbaric oxygen in the treatment of calciphylaxis: a case series and literature review. Nephrology (Carlton) 2015 Jul;20(7):444-450. [doi: 10.1111/nep.12433] [Medline: 25707425]
43. Kemper TC, de Jong VM, Anema HA, van den Brink A, van Hulst RA. Frostbite of both first digits of the foot treated with delayed hyperbaric oxygen: a case report and review of literature. Undersea Hyperb Med 2014;41(1):65-70. [Medline: 24649719]
44. Magnan MA, Gayet-Ageron A, Louge P, Champly F, Joffre T, Lovis C, et al. Hyperbaric oxygen therapy with Iloprost improves digit salvage in severe frostbite compared to Iloprost alone. Medicina (Kaunas) 2021 Nov 22;57(11):1284 [FREE Full text] [doi: 10.3390/medicina57111284] [Medline: 34833502]
45. Hopf HW, Ueno C, Aslam R, Burnand K, Fife C, Grant L, et al. Guidelines for the treatment of arterial insufficiency ulcers. Wound Repair Regen 2006;14(6):693-710. [doi: 10.1111/j.1524-475X.2006.00177.x] [Medline: 17199834]
46. Figueiredo RS, Cohen EJ, Gomes JA, Rapuano CJ, Laibson PR. Conjunctival autograft for pterygium surgery: how well does it prevent recurrence? Ophthalmic Surg Lasers 1997 Feb;28(2):99-104. [Medline: 9054479]
47. Assaad NN, Chong R, Tat LT, Bennett MH, Coroneo MT. Use of adjuvant hyperbaric oxygen therapy to support limbal conjunctival graft in the management of recurrent pterygium. Cornea 2011 Jan;30(1):7-10. [doi: 10.1097/ICO.0b013e3181da58a4] [Medline: 20847683]
48. Dong C, Chen CL, Wang HB. Monocular vision loss after ear filler injection. Aesthetic Plast Surg 2024 Feb 22 (forthcoming). [doi: 10.1007/s00266-024-03903-8] [Medline: 38388798]

49. Jalilian F, Hetz SP, Bostwick J, Boet S. Hyperbaric oxygen therapy for treatment of a late presenting ischaemic complication from hyaluronic acid cosmetic filler injection. BMJ Case Rep 2022 Jul 07;15(7):e249190. [doi: 10.1136/bcr-2022-249190] [Medline: 35798493]
50. Simman R, Bach K. Role of hyperbaric oxygen therapy in cosmetic and reconstructive surgery in ischemic soft tissue wounds: a case series. Eplasty 2022 Dec 07;22:e61 [FREE Full text] [Medline: 36545638]
51. Uittenbogaard D, Lansdorp CA, Bauland CG, Boonstra O. Hyperbaric oxygen therapy for dermal ischemia after dermal filler injection with calcium hydroxylapatite: a case report. Undersea Hyperb Med 2019;46(2):207-210. [Medline: 31051067]
52. Urdiales-Gálvez F, Delgado NE, Figueiredo V, Lajo-Plaza JV, Mira M, Moreno A, et al. Treatment of soft tissue filler complications: expert consensus recommendations. Aesthetic Plast Surg 2018 Apr;42(2):498-510 [FREE Full text] [doi: 10.1007/s00266-017-1063-0] [Medline: 29305643]
53. Hoit G, Bonyun M, Nauth A. Hardware considerations in infection and nonunion management: when and how to revise the fixation. OTA Int 2020 Mar 23;3(1):e055 [FREE Full text] [doi: 10.1097/OI9.0000000000000055] [Medline: 33937680]
54. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. Nat Med 2021 Apr;27(4):601-615 [FREE Full text] [doi: 10.1038/s41591-021-01283-z] [Medline: 33753937]
55. Zilberman-Itskovich S, Catalogna M, Sasson E, Elman-Shina K, Hadanny A, Lang E, et al. Hyperbaric oxygen therapy improves neurocognitive functions and symptoms of post-COVID condition: randomized controlled trial. Sci Rep 2022 Jul 12;12(1):11252 [FREE Full text] [doi: 10.1038/s41598-022-15565-0] [Medline: 35821512]
56. Schouten WR, Briel JW, Auwerda JJ, De Graaf EJ. Ischaemic nature of anal fissure. Br J Surg 1996 Jan;83(1):63-65. [doi: 10.1002/bjs.1800830120] [Medline: 8653368]


## Abbreviations

**ATA:** atmospheres absolute
**ESRD:** end-stage renal disease
**HBO2:** hyperbaric oxygen
**IBD:** inflammatory bowel disease
**IRB:** institutional review board
**NSI:** Neurobehavioral Symptom Inventory
**PCC:** post–COVID-19 condition
**QOL:** quality of life
**RECORD:** Reporting Of Studies Conducted Using Observational Routinely Collected Health Data
**REDCap:** Research Electronic Data Capture
**STROBE:** Strengthening the Reporting of Observational Studies in Epidemiology
**UC:** ulcerative colitis
**UHMS:** Undersea and Hyperbaric Medical Society
**VAS:** visual analog scale


*Edited by T de Azevedo Cardoso, G Eysenbach; submitted 20.10.23; peer-reviewed by E Mehraeen, H Matsui; comments to author 17.01.24; revised version received 17.02.24; accepted 30.07.24; published 20.08.24.*

*Please cite as:*
*Tanaka HL, Rees JR, Zhang Z, Ptak JA, Hannigan PM, Silverman EM, Peacock JL, Buckey JC, Multicenter Registry for Hyperbaric Oxygen Treatment Consortium*
*Emerging Indications for Hyperbaric Oxygen Treatment: Registry Cohort Study*
*Interact J Med Res 2024;13:e53821*
*URL: https://www.i-jmr.org/2024/1/e53821*
*doi:10.2196/53821*
*PMID:39078624*

Original Paper

# Telemedicine Research Trends in 2001-2022 and Research Cooperation Between China and Other Countries Before and After the COVID-19 Pandemic: Bibliometric Analysis

Hanlin Feng[1], MSc; Karin Kurata[2], MSc; Jianfei Cao[3], PhD; Kageyama Itsuki[4], PhD; Makoto Niwa[5], PhD; Atsushi Aoyama[1], PhD; Kota Kodama[4], PhD, RPh

[1]Graduate School of Technology Management, Ritsumeikan University, Ibaraki, Japan

[2]Course of Information Systems Engineering, National Institute of Technology, Tsuruoka College, Tsuruoka, Japan

[3]MergeSystem Inc, Fukuoka, Japan

[4]Medical Data Science Lab, Hoshi University, Tokyo, Japan

[5]Graduate School of Technology, Ritsumeikan University, Ibaraki, Japan

**Corresponding Author:**
Kota Kodama, PhD, RPh
Medical Data Science Lab
Hoshi University
2-4-41 Ebara
Shinagawa-Ku
Tokyo, 142-8501
Japan
Phone: 81 337861011
Email: kodama.kota@hoshi.ac.jp

## Abstract

**Background:** Advancements in technology have overcome geographical barriers, making telemedicine, which offers remote emergency services, healthcare, and medication guidance, increasingly popular. COVID-19 restrictions amplified its global importance by bridging distances.

**Objective:** This study aimed to analyze Chinese and global literature data, present new global telemedicine research trends, and clarify the development potential, collaborations, and deficiencies in China's telemedicine research.

**Methods:** We conducted bibliometrics and network analyses on relevant documents from the Web of Science database from 2001 to 2022. Data collection was completed on October 30, 2023. Considering COVID-19's impact, 2020 was used as a baseline, dividing the data into 2 periods: 2001-2019 and 2020-2022. The development potential was determined based on publication trends. An international coauthorship network analysis identified collaboration statuses and potential. Co-occurrence analysis was conducted for China and the world.

**Results:** We identified 25,333 telemedicine-related research papers published between 2001 and 2022, with a substantial increase during the COVID-19 period (2020-2022), particularly in China (1.93-fold increase), moving its global publication rank from tenth to sixth. The United States, the United Kingdom, and Australia contributed 62.96% of the literature, far ahead of China's 3.90%. Globally, telemedicine research increased significantly post-2020. Between 2001 and 2019, the United States and Australia were central in coauthor networks; post-2020, the United States remained the largest node. Network hubs included the United States, the United Kingdom, Australia, and Canada. Keyword co-occurrence analysis revealed 5 global clusters from 2001 to 2019 (system technology, health care applications, mobile health, mental health, and electronic health) and 2020 to 2022 (COVID-19, children's mental health, artificial intelligence, digital health, and rehabilitation of middle-aged and older adults). In China, the research trends aligned with global patterns, with rapid growth post-2020. From 2001 to 2019, China cooperated closely with Indonesia, India, Japan, Taiwan, and South Korea. From 2020 to 2022, cooperation expanded to Japan, Singapore, Malaysia, and South Korea, as well as Saudi Arabia, Egypt, South Africa, Ghana, Lebanon, and other African and Middle Eastern countries. Chinese keyword co-occurrence analysis showed focus areas in system technology, health care applications, mobile health, big data analysis, and electronic health (2001-2019) and COVID-19, artificial intelligence, digital health, and mental health (2020-2022). Although psychology research increased, studies on children's mental health and middle-aged and older adults' rehabilitation were limited.

**Conclusions:** We identified the latest trends in telemedicine research, demonstrating its significant potential in China and providing directions for future development and collaborations in telemedicine research.

***(Interact J Med Res 2024;13:e40801)*** doi:10.2196/40801

**KEYWORDS**

telemedicine; telehealth; coauthorship analysis; network analysis; bibliometric analysis; co-occurrence analysis

## Introduction

Development of personal computers, communication technology, and professional medical technology have led to improvements in medical-related technology. Telemedicine provides health care solutions in difficult-to-access areas. Health care organizations worldwide are becoming increasingly interested in implementing telemedicine technologies to improve care and services [1]. Telemedicine is the delivery of health care and the exchange of health care information across distances [2]. Telemedicine dates back to the mid-to-late 19th century [2]. Furthermore, one of the earliest reports was published in the early 20th century, with electrocardiogram data transmitted over telephone lines. Currently known telemedicine was formed in the 1920s to 1960s and was largely driven by military and space technology sectors [2,3]. Moreover, only a few individuals used off-the-shelf commercial equipment at the time [4]. Examples of early telemedicine technology include the use of televisions to facilitate consultations between specialists in institutions and general practitioners in state psychiatric hospitals [5]. Furthermore, it was also used to provide expert medical guidance to airport medical centers from major teaching hospitals [6]. Continued development of internet-based audio and video communication technologies, coupled with high demand for a convenient and efficient way to receive care, enabled the development of telemedicine applications [7]. With the rapid development of telemedicine equipment and information communication technology, telemedicine has developed rapidly and is used widely around the world as a new mode of medical service [8,9]. In particular, in China, telemedicine has been used as a crucial method by the government to address the inequality of medical resources between urban and rural areas [10]. The development of telemedicine in China began in the 1980s. In 1986, the Guangzhou Ocean Shipping Company conducted a cross-sea consultation for emergency patients on the oceangoing freighter through telegraph, which was the earliest telemedicine activity in China. In 1997, the Jinwei Medical Network in China was officially opened to provide remote, off-site, real-time, and dynamic live television consultations for patients with severe illness. Subsequently, the medical institutions at all levels in China began to explore and develop telemedicine. After years of effort, the development of telemedicine entered its golden age. In 2017, 22 provinces in China established telemedicine platforms covering 13,000 medical institutions, providing teleconsultation, telediagnosis, and remote medical education [10].

Travel was restricted owing to the COVID-19 pandemic starting in 2019 [11]. However, telemedicine is attracting attention worldwide as it can overcome the problems caused by distances. Telemedicine was increasingly used during the COVID-19 pandemic as a tool to reduce the spread of potential diseases and fill gaps in health care services. Telemedicine emerged as a crucial tool during the COVID-19 pandemic, playing a significant role in health care delivery and management. The global health crisis prompted by the pandemic accelerated the adoption of telemedicine, highlighting its importance in ensuring the continuity of medical services and reducing the risk of infection transmission. Studies have shown that telemedicine was particularly valuable in providing care to patients while minimizing physical contact, especially in situations where in-person visits may have posed health risks due to the highly contagious nature of the virus [12]. Telemedicine could serve as a public health tool to reduce hospital burden, provide continuity of care, and support disease surveillance and management. Particularly, China took protective measures and relied on telemedicine as a response to the pandemic [13,14]. Zhang and Ma [15] quantified the differences between online and face-to-face delivery of health services during the COVID-19 pandemic through a discontinuous time series study in Beijing, China. Moreover, they compared the impact of COVID-19 on the primary outcomes of face-to-face (outpatient, emergency visits, and discharge) and online (online inquiry) health care services. The impact of COVID-19 on health care for different diseases was also analyzed. The average monthly outpatient and discharge volumes of 22 public hospitals in Beijing dropped by 36.33% (2020: 1,720,180 cases; 2019: 2,701,790 cases) and 35.75% (2020: 21,600 cases; 2019: 86,770 cases), respectively, compared with those in 2019. In addition, the average monthly online consultation volume increased by 90.06% (2020: 12,050 cases; 2019: 6340 cases) [15]. After the COVID-19 lockdown, patients with nonsevere disease opted for online consultations. Conversely, patients with severe disease chose hospital care.

Simultaneously, studies are increasingly focusing on the further development of telemedicine. Although telemedicine technology has developed rapidly around the world, research on China's specific development in this field, the challenges it faces, and the status and potential of international cooperation is still limited. Therefore, understanding the current trends and hot spots of publication in telemedicine is necessary. Understanding the pandemic’s effect on telemedicine research through Chinese protective measures can play a significant role in promoting the development of telemedicine.

In light of this, understanding telemedicine research trends in China, where demand for telemedicine is high, and research relationships among China and other countries is of particular interest. This study aimed to analyze Chinese and global literature data, present new global trends in telemedicine research, and clarify the development potential of China's

telemedicine research, potential collaborations, and deficiencies in its own telemedicine research.

Bibliometric analysis is a research method that quantitatively analyzes data from scientific literature and publications. Results of a bibliometric analysis can guide future research by identifying the quality of research and major research areas [16,17]. It includes coauthorship analysis to identify the relationships between institutions and countries [16,17]. Co-occurrence analysis is another methodology used in bibliometric analysis. It investigates co-occurrence relationships between keywords in academic literature and reveals their relationships and structures [16,17]. Co-occurrence analysis allows the identification of research hot spots and the main research directions through keyword clustering. It also allows researchers to obtain information on topics of interest [17-19]. It helps researchers understand the trends in academic research and topics, areas of research focus, and evolution of research fields. Keywords are extracted through co-occurrence analysis [18].

Academically relevant literature on telemedicine bibliometric analyses to date includes the analysis by Armfield et al [20] of 17,932 articles published between 1970 and 2013 in the PubMed database. Furthermore, they divided the time period into 1970 to 1995 and 2009 to 2013. The study examined the themes during the early adoption of telemedicine and compared them with more recently published reports, which provided an understanding of the maturity, scope, and perceptions of telemedicine. Furthermore, they found that the focus of the field shifted from technical issues to clinical applications and assessment.

Sikandar et al [1] conducted a bibliometric analysis that extracted data from 2011 to 2020 from the Scopus database. This study analyzed 1401 articles to examine primary authors, journals, research institutions and countries, and articles with a high number of citations. However, this study lacked data from 2001 to 2011, which was not conducive to an overview of the development of telemedicine. They found that the United States, the Netherlands, and the United Kingdom had many publications in telemedicine. Furthermore, the 10 most studied keywords were identified: mobile health (mHealth), telemedicine, internet, eHealth literacy, technology, self-management, digital health, primary health care, mental health, and electronic health records. This study enabled researchers to understand that telemedicine was an emerging discipline as well as an understudied field. However, telemedicine experienced rapid development after 2020 due to the COVID-19 pandemic. Therefore, updating our understanding regarding telemedicine is essential.

Yanga et al [7] used the Web of Science database to obtain data on telemedicine over a 20-year period, from 1993 to 2012, which produced 7960 publications. Bibliometric analysis revealed that, although the total volume of telemedicine literature increased significantly over the past 20 years, publication activity in each country and region changed over time. The results revealed that the number of telemedicine publications per year increased from 10 in 1993 to 996 in 2012. There was annual growth from 1993 to 2000; however, the growth rate declined between 2000 and 2008. Despite this, growth in telemedicine publications continued to decrease and has remained steady since 2009 [7]. Although the United States leads the cumulative number of telemedicine publications, Norway ranked the highest when ranking countries by publications per capita. The study also found that the annual increase in the number of publications was inconsistent between 1993 and 2012. Moreover, neuroscience and nursing were 2 significant subresearch fields in telemedicine research.

Waqas et al [21] conducted a bibliometric analysis and extracted data through the Web of Science from 2010 to 2019, with a focus on English publications. They found that academic research in telemedicine increased significantly over the 10-year period. Furthermore, research on telemedicine was primarily conducted by institutions in high-income countries. In addition, telemedicine research shifted from radiology to specific disciplines, such as telemedicine, teledermatology, and telecare. Disciplines that promoted joint research included public environment and occupational health, psychiatry, pediatrics, health policy and services, nursing, rehabilitation, radiology, pharmacology, surgery, respiratory medicine, neuroscience, obstetrics, and geriatrics. However, after the COVID-19 pandemic, telemedicine research fostered telemedicine development, which implies the need for further analysis.

Lan et al [14] discussed the research direction of telemedicine during COVID-19, clarified the diseases for which telemedicine technology was used, and identified the medical services it provided. As of December 16, 2021, 5224 telemedicine research papers on COVID-19 were retrieved from the PubMed database. The United States had the most published articles on telemedicine. Simultaneously, most research was related to the provision of health care and mental health services. Although the study mainly explained the application of telemedicine during COVID-19, it did not consider the difference in the timeline before and after the COVID-19 pandemic. Hence, there is a need to further analyze the changes in telemedicine research motivated by COVID-19. Telemedicine's popularity and acceptance, as well as the awareness and recognition of it by patients and health care professionals, may have changed before and after the COVID-19 pandemic. Clarifying these changes is important when considering the future direction of telemedicine.

Waqas et al [21] also found that Web of Science indexed 6896 papers from 2010 to 2019, with an overall h-index of 87. Furthermore, each study was cited an average of 10.64 times. In total, 42,381 papers were cited 73,354 times. In addition, both the number of published papers and number of citations showed an increasing trend from 2010 to 2019. Regionally, publication output (in English) was highest in high-income countries: the United States, Australia, the United Kingdom, Canada, and Germany. However, 2 middle-income countries, India and China, also featured in the top 10 for publication output. These results demonstrated the rapid growth in telemedicine, even before the COVID-19 pandemic. We anticipated that the research landscape and implementation of telemedicine infrastructure would likely advance exponentially during and after the COVID-19 pandemic. This was echoed by a recent report from the American Medical Association, which predicted that "$250 billion in care could shift to telehealth in

the wake of COVID-19, boosting research and infrastructure" [22].

Through preliminary research in telemedicine, we found that the relevant literature increased or decreased every year [1,21]. However, the overall trend was rising. Simultaneously, the United States made the largest contribution to the literature [21]. Research hot spots ranged broadly, including early technology, teleconsultation, telepathology, teleradiology, teledermatology, chronic disease care, and home care.

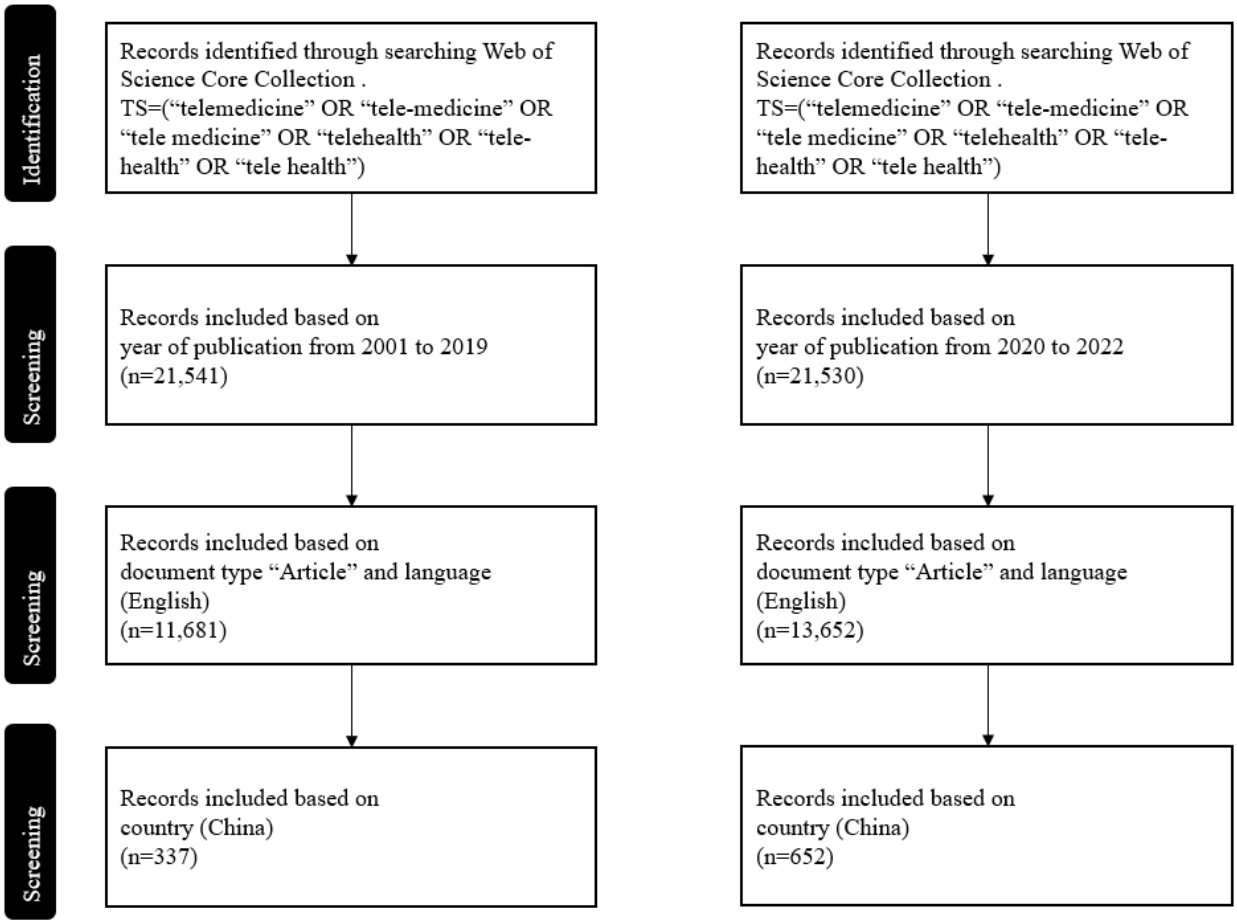
This study aimed to analyze Chinese and global literature data, present new global trends in telemedicine research, and clarify the development potential of China's telemedicine research, potential collaborations, and deficiencies in its own telemedicine research. By identifying and understanding its own opportunities and shortcomings, China could further develop effective strategies to promote the innovation and development of telemedicine technologies and services. Identifying potential countries with which to collaborate will help China establish international collaboration in telemedicine, share knowledge and experience, and jointly respond to global health challenges, such as COVID-19.

## Methods

### Data Collection

We conducted a comprehensive literature search for telemedicine research articles published between 2001 and 2022 in the Web of Science database. A combination of the following keywords was used, with all relevant keywords in 1 search box: topic search = ("telemedicine" OR "tele-medicine" OR "tele medicine") or ("telehealth" OR "tele-health" OR "tele health"). This means that we searched for articles containing any "telemedicine"-related terms and any "telehealth"-related terms at once. In this way, we ensured that all results contained relevant keywords in the fields of telemedicine and telehealth. Our selection criteria were limited to articles published in English within a specified time frame. The search was divided into 2 periods: 2001-2019 and 2020-2022. As illustrated in Figure 1, the initial search yielded 11,681 and 13,652 articles in the first and second periods, respectively. Subsequently, we extracted data that pertained to publications from China, which resulted in 337 and 652 articles in the first and second periods, respectively. Data collection was completed on October 30, 2023.

**Figure 1.** Data collection strategy for the bibliometric analysis of telemedicine research. TS: topic search.




We selected the Web of Science database to extract data for our publication analysis. Since the use and definition of "telemedicine" were not uniformly standardized across the literature and had variations in spelling and terminology, it is critical to encompass such diversity. The term "telemedicine" and its synonymous term "telehealth" were used interchangeably. A previous study highlighted the importance of including both "telemedicine" and "telehealth" in searches for comprehensive representation [7]. This database was chosen for 2 primary reasons. First, it provided extensive citation

coverage across more than 20,000 peer-reviewed journals in diverse fields, such as health, social sciences, and engineering. Second, it was the sole database that enabled citation management across 250 distinct subjects [21]. These features led numerous researchers to consider the Web of Science as an optimal resource for conducting bibliometric analyses [23-25].

Before conducting data analysis, it is crucial to ensure the data's accuracy and reliability. To this end, we performed data cleaning to check for and remove possible duplicate entries in the database, ensuring that each article appeared only once. Databases may list multiple institutions as affiliations for a single author due to the author's personal profile, which may not match the affiliation listed in the article. In such cases, we cleaned the entries to retain only the author's valid affiliation at the time of publication. If the affiliation field for the same author contained multiple institutions, we cleaned the data to reflect only the primary affiliation associated with the publication. This step is critical to preventing misrepresentation of institutional contributions and collaborations. Through these data cleaning steps, we ensured the data set's completeness and accuracy, avoiding bias in the research results due to duplicate or erroneous data, including titles, author names, journal names, and publication dates.

### Use of Software

This study used Microsoft Excel (version 2019), VOSviewer (version 1.6.18), R studio (version 3.3.0), Bibliometrix package (version 4.1.3), Statnet package (version 2019.6), and CiteSpace (version 6.2.R4).

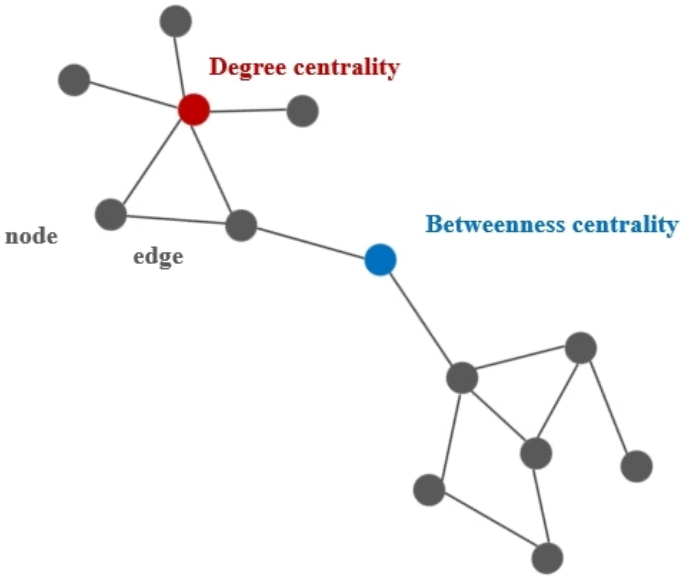
### Data Analysis

We used Microsoft Excel (version 2019) to collect the statistics and organize the following literature data [26,27]: the annual number of publications in China and the world, the top 10 countries in the world in total number of publications from 2001 to 2022, the number of publications in these countries from 2001 to 2019 and 2020 to 2022, and magnification. We also used Microsoft Excel (version 2019) to calculate the ratio of the number of publications in each country to the total number of publications and determine the contribution of each country in this field. Magnification is the publication growth ratio in a period, that is, comparing the ratio of the number of publications in the 2 different time periods, 2001-2019 and 2020-2022. The calculation used the following formula: $N_2/N_1$, where N is the number of publications in each time period [28]. This comparison can be used to measure how quickly the telemedicine field or topic is growing or changing over time. This ratio helps understand whether the field of telemedicine has experienced significant research growth during and after the COVID-19 pandemic.

Furthermore, we performed bibliometric analysis with VOSviewer (version 1.6.18). VOSviewer is a software tool designed to construct and visualize bibliometric networks [17]. Bibliometric analysis is the application of mathematics and statistics [16]. We used VOSviewer and generated bibliometric maps to visualize social networks based on geographical data, which thereby showed international collaborations within the telemedicine domain. Additional analyses were performed via R Studio (version 3.3.0) and incorporated the Bibliometrix and Statnet packages to delineate intercountry relationships in telemedicine research. These packages are tools for bibliometric analysis and scientific mapping, respectively. Bibliometrix is an R package specially designed for bibliometric analysis and provides functions for data import, cleaning, analysis, and bibliometric data visualization [29]. Conversely, Statnet is a software package used to describe and visualize networks of relationships between countries in bibliometric data, with a special focus on international research collaborations [30]. These tools are valuable for researchers who conducted bibliometric studies in various fields. They enable the analysis of publication trends, citation networks, coauthorship patterns, and international collaborations and provide insights into the structure and dynamics of scientific research. By leveraging these packages, researchers gain a comprehensive view of scholarly communications and identify key contributors, influential publications, and emerging trends. The use of Bibliometrix and Statnet in R Studio was consistent with the purpose of conducting bibliometric and network analyses to describe international collaborations in telemedicine. These tools facilitated the visualization and interpretation of social networks for academic collaboration, elucidated the relationships between countries, and disseminated research in telemedicine. We established thresholds, setting the minimum numbers of publications in a country and citations to 10 and 5 times, respectively, to ensure the relevance and importance of network analysis. This approach allowed us to identify China’s role and evolution in international telemedicine collaboration [31,32].

Figure 2 presents a network diagram in which each point is shown as a node within the network structure. Connections between the points, indicative of relationships, are illustrated as edges. Degree centrality is a key measure within this context and was defined as the number of direct neighbors of a node. It was calculated by adding the total number of direct links that a node had with other nodes [33].

**Figure 2.** Description of a network diagram.




Betweenness centrality quantified a node's role as an intermediary within a network. It was measured by the frequency with which a node occurred on the shortest paths between pairs of nodes. In essence, a node with high betweenness centrality acted as a critical bridge within the network's shortest path structure. The calculation of betweenness centrality was derived from the aggregate shortest paths that passed through a given node:

where σst was the total number of shortest paths between nodes s and t and σst(υ) was the number of paths passing through node υ [33,34].

Betweenness centrality is known as an "intermediary agency" role in the network. If one node was in the path to other unique nodes through communicating, connecting, transporting, or trading, then it could be important and is likely to have a high betweenness centrality [35].

We conducted co-occurrence analysis using author-provided keywords in VOSviewer to identify research hot spots in telemedicine. We chose binary counting when using VOSviewer for analysis. This means that each keyword is counted only once, no matter how many times it appears in the document. This is done to ensure that the weight of each keyword in the network graph is not affected by its frequency of occurrence in a single document but is based on its occurrence in multiple documents. The minimum number of co-occurrences was set to 50. This is to ensure that the keywords analyzed are sufficiently representative in international research. For China's data, since the amount of extracted data was lower, the minimum number of co-occurrences was set to 5. This ensures that important keywords are included even in smaller data sets. Due to the search strategy, the keywords telemedicine and telehealth (and keywords with similar meanings) appear more frequently and occupy a larger weight in the co-occurrence network graph. Such keywords are considered to affect the distribution of the remaining keywords; therefore, when mapping the co-occurrence of keywords in China and the world, the keywords used in the search strategy that appeared in the results are deleted, and the results are concentrated on valuable research topic buzzwords [36]. Keywords were standardized, and synonyms and different spellings were merged. For example, we combined synonyms like "COVID19" and "covid-19" into 1 unified keyword.

To examine the changes in hot spots more comprehensively, we used CiteSpace (Basic version 6.2. R4) to visualize the changes over time for international telemedicine research trends [37]. We used CiteSpace (version 6.2. R4) to analyze keyword bursts. CiteSpace is popular scientometric software to analyze and visualize trends, patterns, and frontiers in scientific literature. It includes several key concepts, such as "keyword burst," and has 2 important concepts [38]. Keyword explosions refer to sharp increases in the frequency of the use of certain keywords within a specific period. An explosion could indicate that the represented concept, technology, or topic represented had received widespread attention or become a hot research topic during that period. CiteSpace's (version 6.2.R4) keyword burst analysis helps identify and track research trends and hot areas [38].

## Results

### Telemedicine Research Publications

The publication patterns in telemedicine revealed a global upward trajectory, which was particularly significant over the past 3 years. Although the growth from 2001 to 2019 was gradual, a marked acceleration in publication volume was observed after 2020. The number of telemedicine papers published by the top 10 countries from 2020 to 2022 surpassed those published in the preceding 19 years, as shown in Table 1. This was also notably true for China, where the publication counts from 2020 to 2022 were approximately 1.93 times the

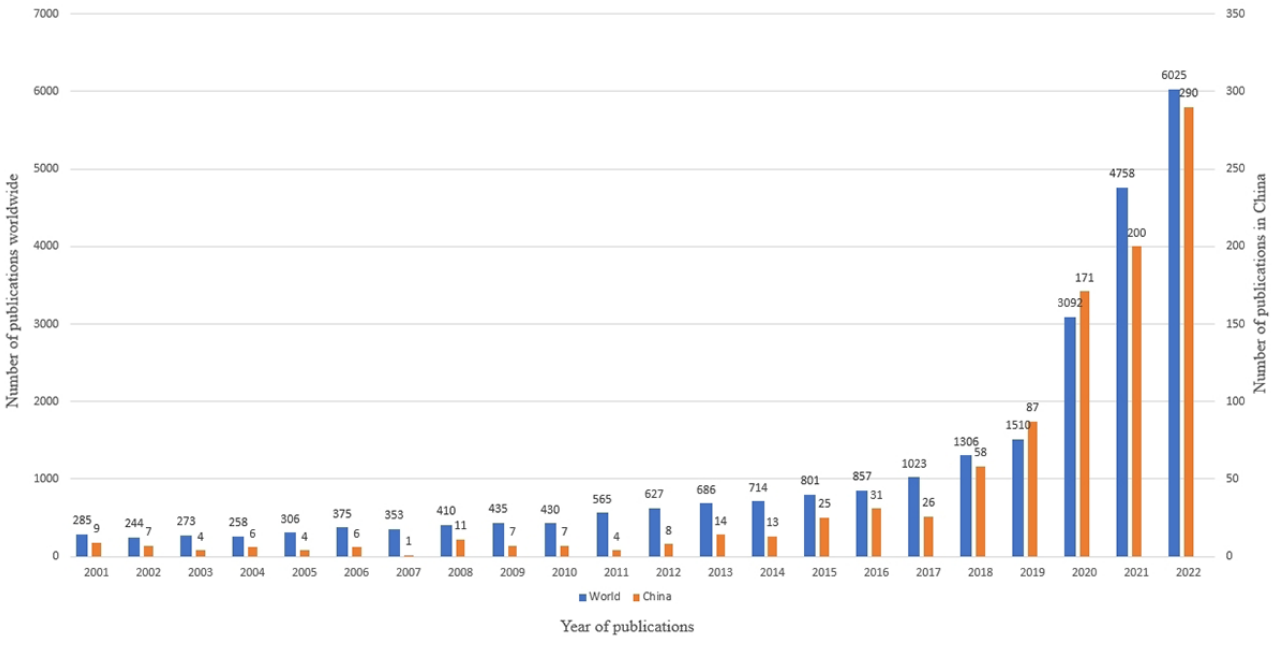
aggregate for those in 2001 to 2019. Table 1 also shows the global ranking of Chinese telemedicine research publications. China's global ranking increased from tenth in 2001 to 2019 to sixth in 2020 to 2022. This surge highlighted the escalating significance of telemedicine, propelled by technological advancements in health care and intensified by the COVID-19 pandemic [20-22]. Figure 3 illustrates the burgeoning global interest in telemedicine research, a trend expected to persist. Figure 3 also charts the trajectory of telemedicine publications in China, which shows a steady increase from 2001 to 2019, followed by a sharp increase in the subsequent years. This pattern mirrored the global trend and highlighted China's intensified focus on telemedicine research in the face of recent health challenges.

**Table 1.** Top 10 countries by publication volume (total number of publications=25,333).

| Rank | Country | Publications in the country, n (%) | Publications by time period | | Magnification[a] |
|---|---|---|---|---|---|
| | | | 2001-2019, n | 2020-2022, n | |
| 1 | United States | 12,112 (47.81) | 5295 | 6817 | 1.29 |
| 2 | Australia | 2035 (8.03) | 1002 | 1033 | 1.03 |
| 3 | England | 1804 (7.12) | 873 | 931 | 1.07 |
| 4 | Canada | 1543 (6.09) | 782 | 761 | 0.97 |
| 5 | Italy | 1201 (4.74) | 485 | 716 | 1.48 |
| 6 | Germany | 1022 (4.03) | 498 | 524 | 1.05 |
| 7 | India | 998 (3.93) | 362 | 636 | 1.76 |
| 8 | China | 989 (3.9) | 337 | 652 | 1.93 |
| 9 | Spain | 842 (3.32) | 401 | 441 | 1.10 |
| 10 | Netherlands | 665 (2.62) | 352 | 313 | 0.89 |

[a]Number of publications before and after 2020.

**Figure 3.** Number of international telemedicine publications.




## International Coauthorship Network for the Telemedicine Research Field

Figures 4A and 4B depict the international coauthorship network in the telemedicine field. The size of each node corresponds to the number of publications in each country, and larger nodes indicate a higher volume. Proximity between 2 nodes suggest the degree of relatedness between the corresponding countries regarding research. An interconnecting line represents collaborative relationships, and thicker lines denote more frequent or substantial collaborative efforts. The different colors assigned by VOSviewer signifies different clusters within the network, where nodes of the same color typically share common properties or characteristics [39]. The network demonstrates the global scale of telemedicine research, showcasing various nations' contributions and cooperative efforts to improve health care through telemedicine technologies. As shown in Figure 4A, the United States and Australia emerged as central hubs for

coauthor networks between 2001 and 2019, which emphasizes their critical roles in telemedicine research collaboration. Although Brazil's node is relatively small, its proximity to the United States and the network's core suggests an important collaborative link. China maintained close cooperative relations with Indonesia, India, Japan, Taiwan, and South Korea. China's main cooperation collaboratives were neighboring Asian countries in the same geographical location. However, network density was more concentrated in Europe and North America, which could reflect more intensive collaboration in these regions.

**Figure 4.** International co-authorship network in telemedicine: (A) 2001-2019 and (B) 2020-2022.

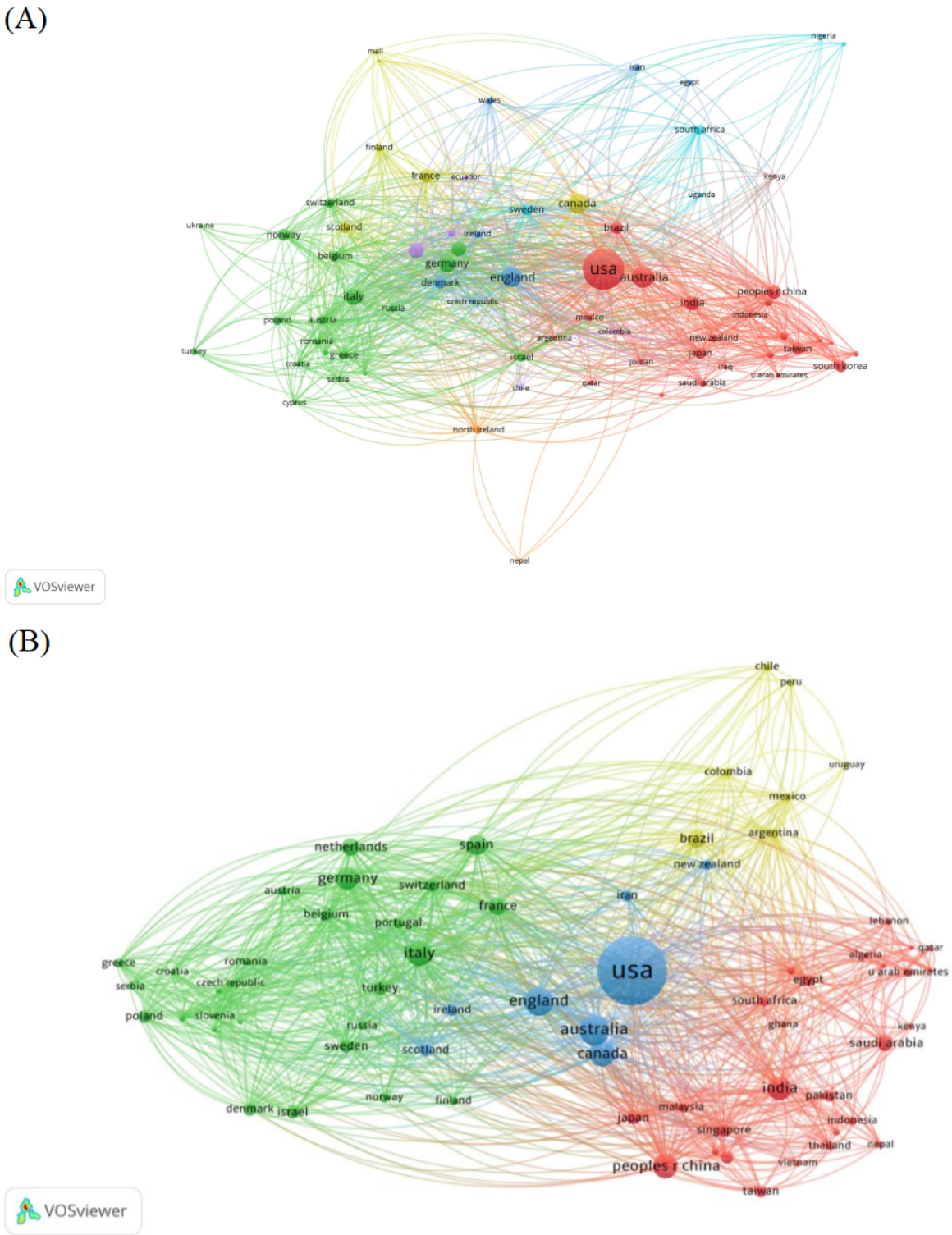



Figure 4B further illustrates the scale of international collaboration in telemedicine between 2020 and 2022. The United States remained the largest node, which represented the most extensive international collaboration. The countries in the center of the network included the United States, the United Kingdom, Australia, and Canada. Importantly, China maintained

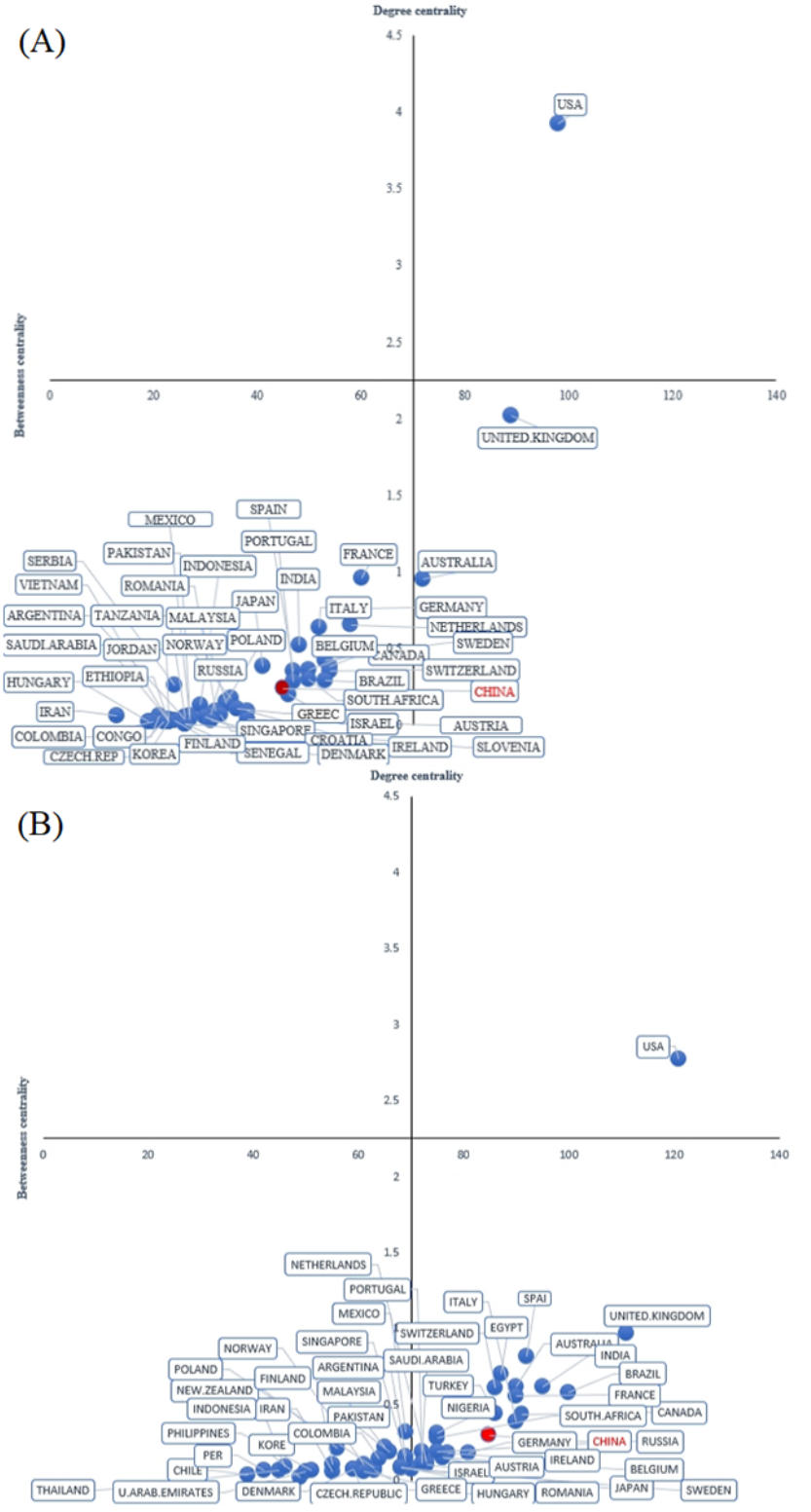
close cooperation with countries from Asia including Japan, Singapore, Malaysia, and South Korea. In addition, China, Saudi Arabia, Egypt, South Africa, Ghana, Lebanon, and other African and Middle Eastern countries unified from the different colored clusters in Figure 4A to a common red cluster in Figure 4B. Cooperation with countries in Africa and the Middle East was also strengthening. Simultaneously, European countries unified from the different colored clusters in Figure 4A to a common green cluster in Figure 4B. Italy is in the central area, and the size of the nodes is relatively large, which indicates more contributions to publications.

This study used RStudio (version 3.3.0) to generate the degree and betweenness centrality for 130 countries in the field of telemedicine. Matrices of the top 50 countries in degree and betweenness centrality from 2001 to 2019 (complete list shown in Multimedia Appendix 1) and 2020 to 2022 (complete list shown in Multimedia Appendix 2) are shown in Figures 5A and 5B, respectively. The central axis divides the matrix into 4 quadrants, with the degree centrality increasing along the X axis from left to right and the betweenness centrality increasing along the Y axis from bottom to top. The X axis quantifies the degree centrality, which reflects the number of connections a country has within the network. In contrast, the Y axis quantifies betweenness centrality, which indicates a country's role as an intermediary within the network.

In Figure 5, the span of the X and Y axes ranged from 0 to 140 and 0 and 4.5, respectively. From 2001 to 2019, shown in Figure 5A, to 2020 to 2022, shown in Figure 5B, the countries as a whole moved to the right and downwards, indicating that countries within the international telemedicine research collaboration network exhibited a general increase in degree and decrease in betweenness centrality. This trend implies that, although the number of international collaborations increased, the role of these countries in the dissemination of research information diminished.

In Figure 5, the United States consistently occupied the quadrant characterized by high degree and betweenness centralities, which indicates its significant role as both a primary collaborator and vital conduit for information flow. In contrast, China was situated in a quadrant of low degree and betweenness centralities from 2001 to 2019. However, from 2020 to 2022, China shifted to the quadrant indicative of high degree centrality but low betweenness centrality, which aligns with the general trend in which international collaborations increased while the centrality in information dissemination decreased [40,41].

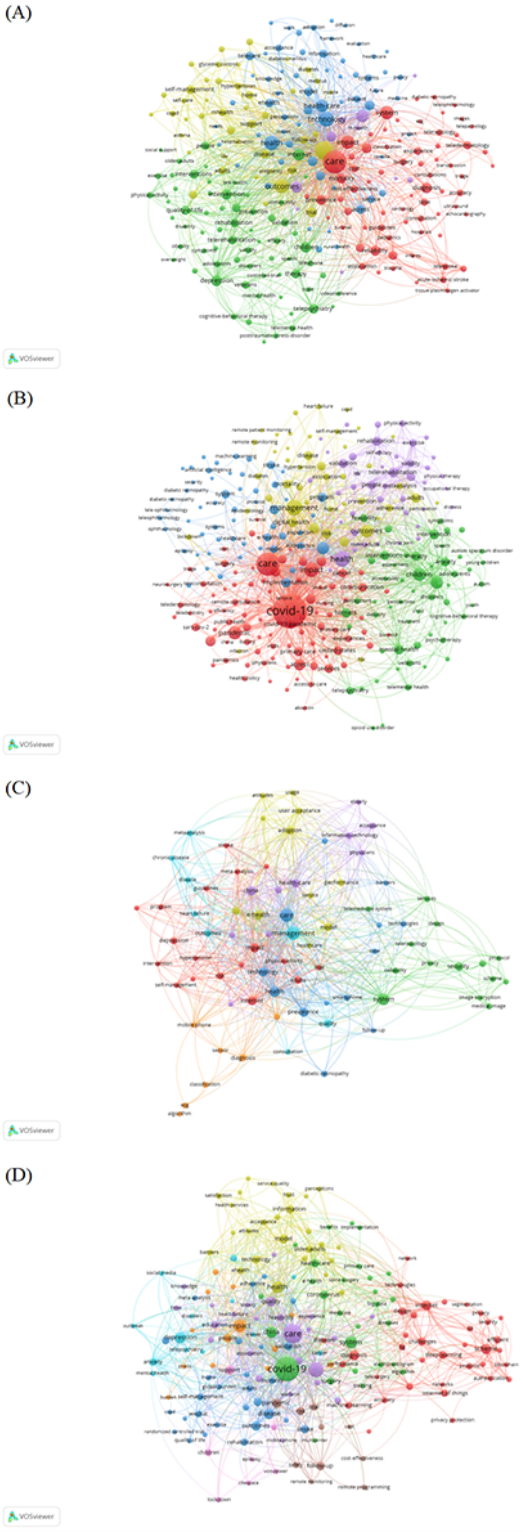
**Figure 5.** Matrix of degree and betweenness centrality in telemedicine: (A) 2001-2019 and (B) 2020-2022.




## Co-Occurrence Analysis of Author Keywords for China and International Trends in the Telemedicine Research Field

Figure 6 presents the keyword co-occurrence network. Figures 6A and 6B represent international trends from 2001 to 2019 and 2020 to 2022, respectively. Figures 6C and 6D represent trends in China from 2001 to 2021 and 2020 to 2022, respectively.

**Figure 6.** Keyword co-occurrence network in telemedicine in (A) the world in 2001-2019, (B) the world in 2020-2022, (C) China in 2001-2019, and (D) China in 2020-2022.




In Figures 6A and 6C (2001-2019), the keyword analysis revealed that “care” and “health” were the predominant nodes, which signifies their foundational role in telemedicine research. The prominence of “technology” underscores its critical influence, while the association with terms, such as “mobile health” and “eHealth,” reflects the importance of digital innovation in telemedicine service provision. The emergence of “user acceptance” and “self-management” as notable keywords denotes the sector's focus on consumer engagement and empowerment of patients in managing their health. The keyword “outcomes” highlights the significance of research results and evaluation of service efficacy in telemedicine. Moreover, the keyword “mobile” suggests an intensive research interest in the application of mobile technologies to telemedicine services. “Health information” and “data” highlight the importance of information management and data analysis, and

the emergence of "quality of care" underscores a commitment to service excellence.

The focus on "security" and "privacy" in Chinese research indicates a growing awareness of the need for data protection measures in telemedicine services. Comparison of Figures 6A and 6C shows that, although international studies were inclined toward nursing and health, China had a unique focus on fundamental technology and systems management, with a special emphasis on health data collection and analysis. This difference could stem from specific telemedicine policies implemented in China [41]. However, compared with Figure 6A, which shows global keyword co-occurrence, Figure 6C shows that China's telemedicine research had fewer keywords in mental health.

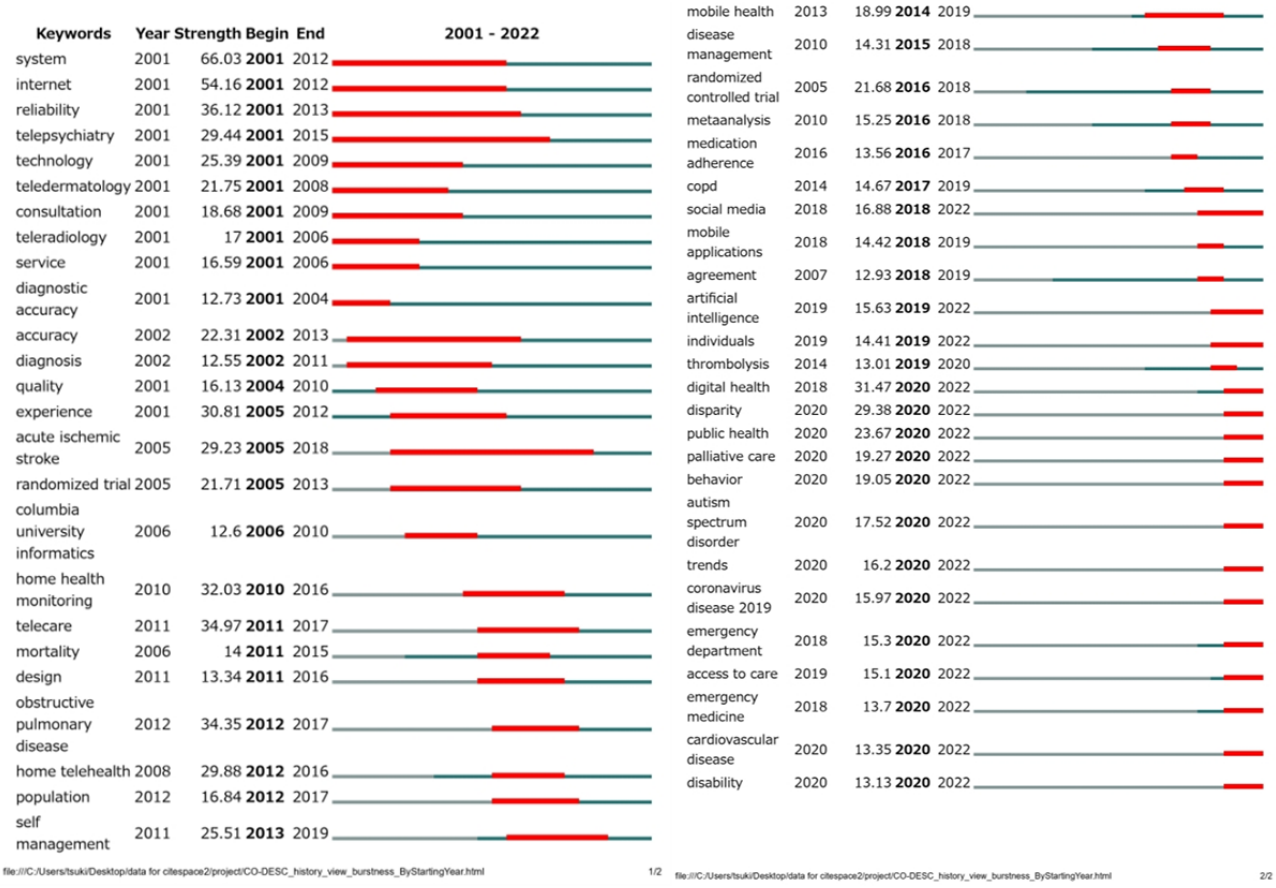
As seen in Figures 6B and 6D, the keyword landscape from 2020 to 2022 was significantly reshaped by the emergence of the COVID-19 pandemic, with the importance of "COVID-19" and "pandemic." This was demonstrated in research discourse. These terms underscored the significant impact of health crises on telemedicine research and implementation. The increase in "mental health" keywords may have relationships with the widespread use of telemedicine in mental health services and psychotherapy. In addition, the mental health issues caused by the pandemic may have provided potential opportunities for telemedicine to implement psychological assistance. The prominent appearance of "data," "analytics," and "artificial intelligence" demonstrates an accelerated convergence of data analytics and artificial intelligence in telemedicine, which points to technological changes triggered by the pandemic. Additionally, consistent references to "telemedicine," "technology," and "digital" reaffirm the centrality of technological innovation in telemedicine services. The emergence of "children" as a keyword highlights the importance of telemedicine for pediatric care. "Patient engagement" and "self-management" suggest that patient autonomy was increasing, which could have been influenced by increased health awareness and limited access to physical care during the pandemic.

Consistent with the global pattern, the term "COVID-19" became the focus of telemedicine research in China during the pandemic. This reflects the significant impact of the pandemic on the health care system and the resulting shift in research focus toward telemedicine. The clustering of terms, such as "eHealth," "mHealth," and "digitalization," highlight the critical role that digital and mobile technologies played in health care delivery in China. The observed increases in the keywords "mental health" and "depression" may indicate an increasing focus in telemedicine research on mental health services and psychotherapy. In addition, the emergence of keywords such as "children" and "rehabilitation" in telemedicine research highlights the growing interest in specialist telemedicine applications. This may reflect an increased awareness of diverse health care needs, including those for the older adult population, pediatric care, and rehabilitation services. These needs may be impacted by the pandemic's broader impact on health care priorities and service delivery models.

Figure 7 shows the bursts of the top 50 global telemedicine keywords between 2001 and 2022. International trends in telemedicine have generally undergone 3 stages of transformation: early focus, mid-term evolution, near-term focus.

Keywords such as "system," "telepsychiatry," and "technology" experienced citation explosions between 2001 and 2012, which reflect the early days of telemedicine, system building, and technology applications in specific fields (eg, telemental health). From 2013 to 2017, there was a rapid increase in citations with keywords such as "telecare," "self-management," and "mobile health," which indicates that telemedicine was beginning to become more personalized as mobile technology became popular and self-management tools developed. This was an indication toward development in the direction of globalization and mobility. Beginning from 2018, there was an explosion of citations for keywords such as "social media," "disparity," and "emergency medicine," possibly due to the increased role of social media in health communication, which heightened concerns of inequalities in medical services and increased demand for remote emergency medical services. Beginning from 2020, pandemic-related keywords, such as "COVID-19," "coronavirus disease," and "emergency department," showed significant citation explosions, which highlights the profound impact of the new coronavirus pandemic on telemedicine research and practice. Telemedicine's focus underwent drastic changes from the initial technical foundation and system construction to the integration of mobile and self-management tools and application of social media during the pandemic.

**Figure 7.** Top 50 international trends with keywords from the co-occurrence network in telemedicine.




## Discussion

### Principal Findings

This study revealed significant insights into the evolution and status of telemedicine research globally, with a special focus on China. The primary findings indicate a substantial increase in telemedicine-related publications over the years, particularly noting a surge from 2020 to 2022, which aligns with the COVID-19 pandemic. This increase reflects the growing global and regional recognition of telemedicine as a critical component of health care systems, especially during times of crisis. We identified a total of 25,333 telemedicine-related research papers published between 2001 and 2022, with the United States, the United Kingdom, and Australia leading in publication numbers. However, China's contribution and global ranking in telemedicine research significantly improved after 2020, showcasing the country's rapidly growing interest and efforts in this field.

### COVID-19 and Telemedicine Research

Overall, the development of communication and professional medical technologies made it easier for many people to access medical services [42]. Both international and Chinese trends steadily increased from 2001, according to literature on telemedicine. However, after 2010, the rapid development of smartphones and mobile technologies enabled the further development of telemedicine research [43]. The COVID-19 pandemic significantly influenced and increased the amount of telemedicine research and number of publications from 2020 to 2022. The COVID-19 pandemic resulted in a significant increase in telemedicine use across medical specialties and geographic locations [44]. Telehealth services expanded rapidly, with a significant shift toward virtual care. This shift was well received by patients and health care providers and had high patient satisfaction scores [45]. Additionally, the pandemic accelerated the adoption of telemedicine in clinical practice across various medical disciplines, such as dentistry, pediatric care, endocrinology, rheumatology, and sports medicine [46]. The widespread adoption of telemedicine facilitated continuity of care for patients and highlighted the potential for telemedicine to be integrated into health care services long after the pandemic [47]. In addition, telemedicine led to positive outcomes, such as reduced hospitalizations and emergency department visits, which demonstrated its effectiveness in remotely managing various medical conditions [48].

### Potential Possibilities in Telemedicine Research

China’s top researchers in telemedicine were actively engaged in research in various medical fields. Specifically, research conducted in western China demonstrated the feasibility, acceptability, and effectiveness of telemedicine during the COVID-19 pandemic. This highlighted the significant contribution of researchers in this region toward the advancement of telemedicine [49]. In addition, pilot studies focused on the evaluation of the utilization rate and cost-effectiveness of telemedicine programs in western China, which indicated that researchers were actively involved in the implementation and evaluation of telemedicine programs [50]. Simultaneously, research on the application of telemedicine in

Gansu Province, China, emphasized that developed cities, such as Beijing, Shanghai, and Guangzhou, had telemedicine consultation centers and successfully managed diseases, such as hypertension and diabetes [51]. Xu et al [52] examined regional heterogeneity in the application and effect of telemedicine in rural primary care centers in China and revealed the diverse implementation and impacts of telemedicine in different regions.

Overall, research by Chinese scholars in telemedicine spanned across many regions, which included western China, Gansu Province, and rural areas. This demonstrated comprehensive and extensive participation in advancing telemedicine practices and technologies.

From a Chinese perspective, Figure 3 shows a significant increase in the number of Chinese publications in telemedicine research over time, especially after 2020, which was consistent with global trends. This growth trend demonstrates China's increasing interest and investment in telemedicine research, which reflects active research activities and potential research momentum. Simultaneously, Table 1 shows China's global ranking in telemedicine research publications. China's global ranking is rising, which shows that its research is valued both domestically and internationally. Moreover, among the multiplier effects of publications in the top 10 countries in Table 1, China's multiplier effect was greater than that of the other countries, which further emphasizes its research potential.

### International Coauthorship in Telemedicine

From 2001 to 2022, the United States was a central node in telemedicine research and had extensive cooperative relationships with other countries. During this period, China gradually developed from a lower degree centrality and lower betweenness centrality to a lower degree centrality and higher betweenness centrality. This indicates that its role as an information dissemination medium in the international collaboration network was small; however, the number of cooperating countries increased. International influence was improved to a certain extent. South American countries (mainly Brazil), western European countries (mainly Italy), and African countries (mainly Egypt and countries in west Africa) all grew in influence. China is currently conducting relatively close cooperative research with Asian and African countries, led by Egypt and countries in west Africa. However, joint research with South American countries (mainly Brazil) and western European countries (mainly Italy) remains relatively slow. These countries are potential research collaborators. Simultaneously, the United States, Australia, the United Kingdom, and Canada were the most influential countries and maintained close cooperation.

### Research Hot Spots in Telemedicine

Between 2001 and 2019, "care" and "health" became the largest research nodes in telemedicine literature, which reflected that health care was the core of telemedicine research. Technology played a significant role in telemedicine, particularly in mHealth and eHealth applications. Telemedicine research in China focused on the management and application of technology. User acceptance, mobile technology, health information, and data analytics were key points in Chinese studies, which reflects China's special emphasis on telemedicine policy and data management. From 2020 to 2022, "COVID-19" and "pandemic" become prominent keywords in global telemedicine research, which indicates that the pandemic greatly affected telemedicine research and practice. This forced researchers to focus on various areas, such as pandemic management, public health, medical disparities, and digital health. Mental health, urgent care, and data analytics were also areas of focus, which indicates the impact of the pandemic on the demand for telehealth services and technology. Simultaneously, the emergence of keywords shown in Figure 7 reflects the forefront of the research field. The emergence of some keywords, such as "big data" and "artificial intelligence," indicates that telemedicine research was beginning to integrate advanced technologies, which could potentially completely change the field.

Simultaneously, the world is working on telemedicine for all ages. However, as far as Chinese keywords were concerned, relatively few studies focused on telerehabilitation for children and older adults. In addition, few studies examined telemental health for children. China should also consider strengthening telemedicine research across all age groups. The world's latest telemedicine research direction is toward the development of artificial intelligence, digital medicine, and individualization. China has also conducted extensive research in these fields. In the future, we will continue to deepen research on artificial intelligence and digital medicine in telemedicine.

### Future Prospects for Research in Telemedicine

This study demonstrated China's significant development in the research field of telemedicine. However, betweenness centrality remained low, which indicates the need for Chinese researchers to seek opportunities that allow them to be the mediators while bonding in smaller subnetworks.

Through this analysis, we found that telemedicine research is likely to focus on improving service efficiency and quality with the advancement of technology, especially in artificial intelligence and mHealth. Future telemedicine research should explore the application and integration of these technologies in the medical field. Personalized medicine and increased patient engagement are expected to be crucial. As international collaboration increases, future telemedicine research should place greater emphasis on knowledge sharing and collaboration internationally. Special keywords, such as "lockdown" and "tracing," were revealed in the co-occurrence analysis for China, which suggests that telemedicine research also targeted specific regional, social, and cultural factors. This demonstrates that solutions must be identified for each culture, region, and country when it comes to mitigating social problems and health challenges through telemedicine development. Policy and regulation will remain an important topic in telemedicine research, particularly regarding data security and privacy protection.

Simultaneously, the impact of the COVID-19 pandemic on telemedicine will be reflected in future research, including the exploration of public health emergency management, vaccination, pandemic monitoring, and remote patient management. Future research is expected to explore how

telemedicine technology can help respond to global health crises. Different cultural and social contexts may bring different acceptance and needs for telemedicine. Considering how to provide medical services that are more adaptable to different cultural and social environments will be a direction for future research.

### Limitations

This study has 4 limitations. First, we only collected bibliometric data from the Web of Science database and excluded other search engines, such as Scopus, PubMed, IEEE, and Google Scholar. Second, we extracted only articles published in English. Third, although we aimed to identify research trends in telemedicine in China, we obtained limited articles on telemedicine, as we only used those published in Web of Science. Fourth, some data in the literature data collected through Web of Science could be changed due to electronic publishing or document editing and publication time. Hence, this may impact the data analysis results.

### Conclusions

This study provides the latest trends in telemedicine research, demonstrates that telemedicine research has considerable potential in China, and provides directions for future development. Simultaneously, China's future remote medical research collaboratives and research areas are also of reference value. Although this study shows significant growth in telemedicine research in China, it points out that China's status as a center for international collaboration is still low. In the future, telemedicine research is likely to focus on improving service efficiency and quality, particularly by leveraging technological advances in artificial intelligence and mHealth. In addition, personalized medicine and increased patient engagement will be important trends in the future.

## Acknowledgments

We would like to express our gratitude to Ritsumeikan University for providing access to the Web of Science database. We also thank all the study participants for their constructive suggestions and guidance. This work was supported by the Japan-Ritsumeikan University Ritsumeikan Advanced Research Academy (RARA) Student Fukuro Scholarship. The authors sincerely thank the Japan-Ritsumeikan University RARA Student Fukuro Scholarship for its generous support and assistance.

## Conflicts of Interest

KK has received research donations for Ritsumeikan University from Kumagaya Gumi, Sumitomo Life Insurance, Merge system, and Pasona. All other authors declare no conflicts of interest.

Multimedia Appendix 1
List of degree and betweenness centrality by country from 2001 to 2019.
[XLSX File (Microsoft Excel File), 13 KB - ijmr_v13i1e40801_app1.xlsx ]

Multimedia Appendix 2
List of degree and betweenness centrality by country from 2020 to 2022.
[XLSX File (Microsoft Excel File), 13 KB - ijmr_v13i1e40801_app2.xlsx ]

## References

1. Sikandar H, Vaicondam Y, Parveen S, Khan N, Qureshi MI. Bibliometric analysis of telemedicine and e-Health literature. Int. J. Onl. Eng 2021 Nov 29;17(12):52-69. [doi: 10.3991/ijoe.v17i12.25483]
2. Craig J, Patterson V. Introduction to the practice of telemedicine. j telemed telecare 2005 Jan 01;11(1):3-9. [doi: 10.1258/1357633053430494] [Medline: 15829036]
3. Currell R, Urquhart C, Wainwright P, Lewis R. Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2000(2):CD002098. [doi: 10.1002/14651858.CD002098] [Medline: 10796678]
4. Einthoven W. The telecardiogram. American Heart Journal 1957;53(4):602-615 [FREE Full text] [doi: 10.1016/0002-8703(57)90367-8]
5. Benschoter RA, Eaton MT, Smith P. Use of videotape to provide individual instruction in techniques of psychotherapy. J Med Educ 1965 Dec;40(12):1159-1161. [doi: 10.1097/00001888-196512000-00006] [Medline: 5839291]
6. Dwyer TF. Telepsychiatry: psychiatric consultation by interactive television. Am J Psychiatry 1973 Aug;130(8):865-869. [doi: 10.1176/ajp.130.8.865] [Medline: 4716685]
7. Yang Y, Iqbal U, Ching JH, Ting JB, Chiu H, Tamashiro H, et al. Trends in the growth of literature of telemedicine: A bibliometric analysis. Comput Methods Programs Biomed 2015 Dec;122(3):471-479. [doi: 10.1016/j.cmpb.2015.09.008] [Medline: 26415760]
8. Tsirintani M. Strategic procedures and revisions for implementing telemedicine and telecare in Greece. Appl Clin Inform 2012;3(1):14-23 [FREE Full text] [doi: 10.4338/ACI-2011-08-R-0048] [Medline: 23616897]

9. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Access and quality of care in direct-to-consumer telemedicine. Telemed J E Health 2016 Apr;22(4):282-287 [FREE Full text] [doi: 10.1089/tmj.2015.0079] [Medline: 26488151]
10. Cui F, Ma Q, He X, Zhai Y, Zhao J, Chen B, et al. Implementation and application of telemedicine in China: cross-sectional study. JMIR Mhealth Uhealth 2020 Oct 23;8(10):e18426 [FREE Full text] [doi: 10.2196/18426] [Medline: 33095175]
11. Wang Y, Yang J, Ma H, Dong X, Xie G, Ye S, et al. Application of telemedicine in the COVID-19 epidemic: An analysis of Gansu Province in China. PLoS One 2021;16(8):e0249872 [FREE Full text] [doi: 10.1371/journal.pone.0249872] [Medline: 34347779]
12. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
13. AlMannai A, AlShaikh R, Hassan H. The clinical effectiveness and patient satisfaction of teleconsultation during the period of COVID19 pandemic in the ENT department at KHUH. Bahrain Medical Bulletin 2021;43(3):556-559 [FREE Full text]
14. Lan X, Yu H, Cui L. Application of telemedicine in COVID-19: a bibliometric analysis. Front Public Health 2022 May 26;10:908756 [FREE Full text] [doi: 10.3389/fpubh.2022.908756] [Medline: 35719666]
15. Zhang S, Ma C. How has the COVID-19 pandemic affected the utilisation of online consultation and face-to-face medical treatment? An interrupted time-series study in Beijing, China. BMJ Open 2023 Feb 10;13(2):e062272 [FREE Full text] [doi: 10.1136/bmjopen-2022-062272] [Medline: 36764719]
16. Caglar C, Demir E, Kucukler FK, Durmus M. A bibliometric analysis of academic publication on diabetic retinopathy disease trends during 1980-2014: a global and medical view. Int J Ophthalmol 2016;9(11):1663-1668 [FREE Full text] [doi: 10.18240/ijo.2016.11.21] [Medline: 27990373]
17. Soytas RB. A bibliometric analysis of publications on COVID-19 and older adults. Ann Geriatr Med Res 2021 Sep;25(3):197-203 [FREE Full text] [doi: 10.4235/agmr.21.0060] [Medline: 34229370]
18. Sedighi M. Application of word co-occurrence analysis method in mapping of the scientific fields (case study: the field of Informetrics). Library Review 2016;65(1/2):52-64. [doi: 10.1108/LR-07-2015-0075]
19. Cao J, Lim Y, Sengoku S, Guo X, Kodama K. Exploring the shift in international trends in mobile health research from 2000 to 2020: bibliometric analysis. JMIR Mhealth Uhealth 2021 Sep 08;9(9):e31097 [FREE Full text] [doi: 10.2196/31097] [Medline: 34494968]
20. Armfield NR, Edirippulige S, Caffery LJ, Bradford NK, Grey JW, Smith AC. Telemedicine--a bibliometric and content analysis of 17,932 publication records. Int J Med Inform 2014 Oct;83(10):715-725. [doi: 10.1016/j.ijmedinf.2014.07.001] [Medline: 25066950]
21. Waqas A, Teoh SH, Lapão LV, Messina LA, Correia JC. Harnessing telemedicine for the provision of health care: bibliometric and scientometric analysis. J Med Internet Res 2020 Oct 02;22(10):e18835 [FREE Full text] [doi: 10.2196/18835] [Medline: 33006571]
22. Henry TA. After COVID-19, $250 billion in care could shift to telehealth. American Medical Association. 2020 Jun 18. URL: https://www.ama-assn.org/practice-management/digital/after-covid-19-250-billion-care-could-shift-telehealth [accessed 2023-10-03]
23. Uribe-Toril J, Ruiz-Real JL, Nievas-Soriano BJ. A study of eHealth from the perspective of social sciences. Healthcare (Basel) 2021 Jan 21;9(2):108 [FREE Full text] [doi: 10.3390/healthcare9020108] [Medline: 33494182]
24. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical activity, sedentary behavior, and diet-related eHealth and mHealth research: bibliometric analysis. J Med Internet Res 2018 Apr 18;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]
25. Wang Z, Deng Z, Wu X. Status quo of professional patient relations in the internet era: bibliometric and co-word analyses. Int J Environ Res Public Health 2019 Apr 02;16(7):1183 [FREE Full text] [doi: 10.3390/ijerph16071183] [Medline: 30986980]
26. Turner A, Brazier J, Bishop C, Chavda S, Cree J, Read P. Data analysis for strength and conditioning coaches: using Excel to analyze reliability, differences, and relationships. Strength and Conditioning Journal 2015;37(1):76-83. [doi: 10.1519/SSC.0000000000000113]
27. Liu F, Lei G, Jia X, He S, Dang Y. Worldwide trends of forensic dentistry: a 20-year bibliometric analysis in Pubmed. Rom J Leg Med 2016 Oct;24(3):236-241. [doi: 10.4323/RJLM.2016.236]
28. Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Asso for Info Science & Tech 2015 Apr 29;66(11):2215-2222. [doi: 10.1002/asi.23329]
29. Derviş H. Bibliometric analysis using Bibliometrix an R Package. JSCIRES 2020 Jan 06;8(3):156-160. [doi: 10.5530/jscires.8.3.32]
30. Aria M, Cuccurullo C. Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics 2017 Nov;11(4):959-975. [doi: 10.1016/j.joi.2017.08.007]
31. Yashiro K, Lim Y, Sengoku S, Kodama K. Recent trends in interorganizational deal networks in pharmaceutical and biotechnology industries. Drug Discov Today 2023 Mar;28(3):103483. [doi: 10.1016/j.drudis.2022.103483] [Medline: 36584874]

32. Dias A, Ruthes S, Lima L, Campra E, Silva M, Bragança de Sousa M, et al. Network centrality analysis in management and accounting sciences. RAUSP 2019 Dec 09;55(2):207-226. [doi: 10.1108/rausp-02-2019-0021]
33. Zhang J, Luo Y. Degree centrality, betweenness centrality, and closeness centrality in social network. Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017) 2017:300-303. [doi: 10.2991/MSAM-17.2017.68]
34. Brandes U. A faster algorithm for betweenness centrality. The Journal of Mathematical Sociology 2001 Jun;25(2):163-177. [doi: 10.1080/0022250x.2001.9990249]
35. Freeman LC. A set of measures of centrality based on betweenness. Sociometry 1977 Mar;40(1):35. [doi: 10.2307/3033543]
36. Madjido M, Espressivo A, Maula AW, Fuad A, Hasanbasri M. Health information system research situation in Indonesia: a bibliometric analysis. Procedia Computer Science 2019;161:781-787. [doi: 10.1016/j.procs.2019.11.183]
37. Chen C, Ibekwe‐SanJuan F, Hou J. The structure and dynamics of cocitation clusters: A multiple‐perspective cocitation analysis. J. Am. Soc. Inf. Sci 2010 Jun 14;61(7):1386-1409. [doi: 10.1002/asi.21309]
38. Sabe M, Pillinger T, Kaiser S, Chen C, Taipale H, Tanskanen A, et al. Half a century of research on antipsychotics and schizophrenia: A scientometric study of hotspots, nodes, bursts, and trends. Neurosci Biobehav Rev 2022 May;136:104608 [FREE Full text] [doi: 10.1016/j.neubiorev.2022.104608] [Medline: 35303594]
39. van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics 2017 Feb 27;111(2):1053-1070 [FREE Full text] [doi: 10.1007/s11192-017-2300-7] [Medline: 28490825]
40. Zhao F, Shi B, Liu R, Zhou W, Shi D, Zhang J. Theme trends and knowledge structure on choroidal neovascularization: a quantitative and co-word analysis. BMC Ophthalmol 2018 Apr 03;18(1):86 [FREE Full text] [doi: 10.1186/s12886-018-0752-z] [Medline: 29614994]
41. Gao J, Fan C, Chen B, Fan Z, Li L, Wang L, et al. Telemedicine is becoming an increasingly popular way to resolve the unequal distribution of healthcare resources: evidence from China. Front Public Health 2022;10:916303 [FREE Full text] [doi: 10.3389/fpubh.2022.916303] [Medline: 35874991]
42. Srivastava S, Pant M, Abraham A, Agrawal N. The technological growth in eHealth services. Comput Math Methods Med 2015;2015:894171-894118 [FREE Full text] [doi: 10.1155/2015/894171] [Medline: 26146515]
43. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
44. Betancourt JA, Rosenberg MA, Zevallos A, Brown JR, Mileski M. The impact of COVID-19 on telemedicine utilization across multiple service lines in the United States. Healthcare (Basel) 2020 Oct 01;8(4):380 [FREE Full text] [doi: 10.3390/healthcare8040380] [Medline: 33019667]
45. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. J Med Internet Res 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: 10.2196/20786] [Medline: 32810841]
46. Al Mohaya MA, Almaziad MM, Al-Hamad KA, Mustafa M. Telemedicine among oral medicine practitioners during COVID-19 pandemic and its future impact on the specialty. RMHP 2021 Oct;Volume 14:4369-4378. [doi: 10.2147/rmhp.s325777]
47. Swidan A, Alnoon NA, Makki I, Zidan M, Alhammadi H, Rahmani N, et al. Effect of COVID-19 pandemic on patient utilization of the telemedicine services in Dubai. Dubai Med J 2022 Mar 14;5(2):110-116 [FREE Full text] [doi: 10.1159/000522346]
48. Raskas MD, Feuerstein-Mendik GJ, Gerlacher G, Cohen S, Henning S, Cramer JM, et al. Epidemiology of 30,000 pediatric urgent care telemedicine visits in the era of COVID-19. Telemed J E Health 2022 Oct 01;28(10):1404-1411 [FREE Full text] [doi: 10.1089/tmj.2021.0482] [Medline: 35172122]
49. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: experiences from western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
50. Wang T, Li J, Zhu C, Hong Z, An D, Yang H, et al. Assessment of utilization and cost-effectiveness of telemedicine program in western regions of China: a 12-year study of 249 hospitals across 112 cities. Telemed J E Health 2016 Nov;22(11):909-920. [doi: 10.1089/tmj.2015.0213] [Medline: 27314300]
51. Chen J, Zheng B, Yin L, Zhou Q, Liu W, Li P, et al. Exploring agreement and feasibility between virtual home visits and in-person home visits for peritoneal dialysis patients—a paired study. Renal Failure 2022 Mar 13;44(1):490-502. [doi: 10.1080/0886022x.2022.2049305]
52. Xu W, Pan Z, Lu S, Zhang L. Regional heterogeneity of application and effect of telemedicine in the primary care centres in rural China. Int J Environ Res Public Health 2020 Jun 24;17(12):1 [FREE Full text] [doi: 10.3390/ijerph17124531] [Medline: 32599689]


## Abbreviations

**mHealth:** mobile health

*Edited by T de Azevedo Cardoso; submitted 29.01.24; peer-reviewed by H Guo, I Maniu; comments to author 09.03.24; revised version received 23.03.24; accepted 19.06.24; published 30.08.24.*

*Please cite as:*
*Feng H, Kurata K, Cao J, Itsuki K, Niwa M, Aoyama A, Kodama K*
*Telemedicine Research Trends in 2001-2022 and Research Cooperation Between China and Other Countries Before and After the COVID-19 Pandemic: Bibliometric Analysis*
*Interact J Med Res 2024;13:e40801*
*URL: https://www.i-jmr.org/2024/1/e40801*
*doi:10.2196/40801*
*PMID:*

Original Paper

# Telemedicine for Patients With Systemic Lupus Erythematosus in a Publicly Funded Hospital System: Retrospective Study

Sebastian Bruera[1], MD; Kristen Andrews Staggers[1], MS; Maria Eugenia Suarez-Almazor[2], MD, PhD; Sandeep Krishna Agarwal[1], MD, PhD

[1]Department of Immunology, Allergy, and Rheumatology, Baylor College of Medicine, Houston, TX, United States

[2]Department of Health Services Research, MD Anderson Cancer Center, University of Texas, Houston, TX, United States

**Corresponding Author:**
Sebastian Bruera, MD
Department of Immunology, Allergy, and Rheumatology
Baylor College of Medicine
7200 Cambridge Suite 10b
Houston, TX, 77030
United States
Phone: 1 7138822370
Email: bruera@bcm.edu

## Abstract

**Background:** Systemic lupus erythematosus (SLE) is a chronic autoimmune disease that requires frequent clinic and laboratory visits. However, patients with SLE, particularly those who are underresourced, have unacceptably high rates of no-shows.

**Objective:** This study aims to determine no-show rates associated with telemedicine visits during the COVID-19 pandemic in comparison to no-show rates associated with contemporaneous and historic in-person visits.

**Methods:** We performed a retrospective cohort study in a publicly funded county hospital system in Houston, Texas. We identified a cohort of established patients with SLE by the *International Classification of Diseases* codes that were independently confirmed as SLE by a review of medical records. We identified patients who were seen from March to December in 2018, 2019, and 2020 (to reflect the height of the COVID-19 pandemic and account for seasonal changes in disease activity). Our primary outcome was the percentage of no-shows for rheumatology clinic appointments. Our secondary outcome was laboratory use adherence, which was defined as lupus-specific blood and urine studies conducted within 30 days of the scheduled appointment. Covariates included age, sex, race, ethnicity, and SLE-related prescription drugs.

**Results:** We included 156 patients with SLE in our analysis. Most were female (n=141, 90.4%), were Hispanic (n=75, 49.3%), and had a median age of 43 (range 19-80) years. In 2020, the no-show rate for telemedicine was 5.5% (10/182) compared to a no-show rate of 16.2% (31/191) for in-person visits (*P*=.002). After multivariable adjustment for covariates, the odds of no-show were lower for telemedicine visits (odds ratio 0.39, 95% CI 0.20-0.77). There were no differences in adherence to laboratory testing.

**Conclusions:** Telemedicine visits had decreased odds of no-shows without difference in laboratory testing adherence after adjustment for covariates. More research is needed to determine the clinical impact of telemedicine on patients with SLE.

***(Interact J Med Res 2024;13:e49065)*** doi:10.2196/49065

**KEYWORDS**

lupus; systemic lupus erythematosus; telemedicine; COVID-19; access to care; autoimmune disease; no-show; socioeconomic status; adherence; laboratory test; management

## Introduction

The COVID-19 pandemic triggered widespread and emergent use of telemedicine as an option for patients to avoid exposure to the SARS-CoV-2 virus [1,2]. The use of telemedicine has been especially important in patients with chronic diseases, such as systemic lupus erythematosus (SLE), who are at a high risk of severe COVID-19 and may benefit from less public exposure [3,4]. Missed appointments in general are associated with an increased risk of mortality and adverse outcomes [5-7]. Telemedicine has the potential benefit of improving no-shows by making clinic visits more accessible. This may be particularly

important for patients of lower socioeconomic status (SES) who may have difficulties attending visits because of transportation, work, or financial factors. Yet, it has been suggested that telemedicine services are less likely to be used in populations with lower SES [8,9].

Despite the potential advantages of telemedicine, especially for patients of low SES, there are still important concerns, which should be considered when telemedicine is implemented for patients with SLE. First, patients with SLE are usually followed every 3 to 4 months and require serial evaluations including blood pressure monitoring, physical examination, and clinical laboratory tests—regardless of the presence or absence of symptoms or examination findings [10]. It is unclear how telemedicine may affect adherence to visits and adherence to laboratory testing since patients may be more likely to obtain blood work if they are already at the clinic.

We performed a retrospective cohort study among patients with SLE managed in a county hospital system in Houston, Texas. This population is highly diverse, with most patients being underinsured or uninsured and having low SES. We hypothesized that patients with SLE in this system will have lower no-show rates with telemedicine modalities, such as telephone and video visits, compared to in-person visits. Furthermore, we explored patient characteristics (such as medications and age) and their association with visit types. We also examined the impact of telemedicine on patients' adherence to laboratory testing.

## Methods

### Design and Patient Population

We performed a retrospective analysis of patients with SLE seen in the Harris Health System (HHS). The HHS is a fully integrated health care system that provides care to residents of Harris County, Texas, which has an estimated population of 4.7 million [11]. The patient population at the HHS includes 54% who are uninsured, 22% who have Medicaid, 12% who have Medicare, and 13% who have private insurance. Most uninsured patients qualify for HHS insurance ("the gold card") that provides partial or full reimbursement for care to patients with a household income that does not exceed 150% of the federal poverty level. Once a patient receives a gold card, the copay for any visits is dictated by income stratification and ranges.

### Patient Selection

We identified a cohort of established patients with SLE using the *International Classification of Diseases, 10th Revision* diagnostic codes (M32.x, excluding M32.0). Patients with SLE were included if they were seen at the HHS rheumatology clinics by a rheumatologist at least once between March 2020 and September 2020 (at a time when telemedicine was implemented because of the pandemic). Data were initially collected through the information technology services provided by the HHS. Patients were seen at a large HHS rheumatology teaching clinic staffed by 7 rheumatologists. The diagnosis of SLE was independently confirmed by chart review by a rheumatologist (SB) if they met the American College of Rheumatology 2019 diagnostic criteria. Patients were offered telemedicine visits either by telephone encounters or by video with a secured third-party platform (Doximity). Between March 2020 and September 2020, the HHS rheumatology clinic offered both in-person visits along with telemedicine encounters. The decision to have a telemedicine versus in-person encounter was driven by patient preference.

As controls, we identified cohorts of patients with SLE seen in the HHS rheumatology clinic from March to September of 2018 and 2019 (before the pandemic). We limited the control cohort to patients with SLE seen from March to September to account for potential seasonal changes in practice patterns and disease activity, which have been previously described [12]. We also identified a subgroup for analysis of patients who were seen at least once in both 2019 and 2020. This subgroup analysis of "no-shows" was performed to examine trends for the same patients who had attended at least 1 follow-up appointment each year.

### Outcomes

Our primary outcome was the percentage of no-shows for rheumatology clinic follow-up appointments. No-shows were defined as visits for which patients did not show up or that were canceled by the patient within the same day. Clinic visits rescheduled by patients prior to 24 hours before the clinic appointment were not considered no-shows. Secondary outcomes included laboratory testing such as complete blood count, comprehensive metabolic panel, urinalysis, serum complement levels (C3 and C4), and serum titers of anti–double-stranded DNA (anti-dsDNA) antibodies within 30 days before or after each completed clinic visit. It is the standard in rheumatology clinics that all patients with SLE obtain laboratory testing and have a clinic appointment at least every 3 months, regardless of disease activity [10]. All laboratory studies for patients seen at HHS clinics are done within the HHS at 1 of 17 clinics or 2 large hospitals. Laboratory tests can be ordered as a preclinic laboratory test (performed within 14 days of a clinic appointment) or obtained on the day of their clinic visit. For telemedicine encounters, blood work can be obtained before the clinic or at patients' convenience for any day of their preference at the closest HHS clinical laboratory.

### Covariates

Baseline demographics included age, sex, race, ethnicity, and use of SLE-related prescribed drugs at the initial visit in the period of interest. We included baseline demographics as a covariate due to multiple studies showing differences in digital literacy among patients of different ages, races, and ethnicities [13]. The insurance coverage for each specific visit appointment was not available, as insurance status can change over time; however, as previously mentioned, over 85% of patients that are seen in our clinics are uninsured or publicly insured.

We also included whether patients were prescribed SLE-specific drugs (hydroxychloroquine, mycophenolate, azathioprine, methotrexate, rituximab, belimumab, tacrolimus, prednisone, and cyclophosphamide) in 2018, 2019, or 2020. Our data included medications prescribed by providers at each visit but did not include whether patients had filled prescriptions (ie, could not measure adherence). Some patients may have a

6-month active prescription for a drug that may not necessarily be refilled at a 3-month follow-up visit. Due to this, we included SLE-specific drug prescriptions as a variable of whether the patient was ever (at any 1 time point) prescribed (or refilled) a medication throughout the year (2018, 2019, or 2020) as opposed to by visit. We also included codes for infusions of rituximab, belimumab, and cyclophosphamide. We assumed that drug therapies may be an important covariate as some drugs, for instance, mycophenolate mofetil, require more frequent laboratory monitoring than others, such as hydroxychloroquine. We did not use drug therapy as a surrogate for disease activity.

We also used our covariates to determine any associations with visit types in 2020 when telemedicine was more readily available. We divided patients into either all in-person visits, 2 or more telemedicine visits, or in-person visit with 1 telemedicine visit to determine if there were differences between the covariates.

### Statistical Analysis

Patient and visit characteristics were summarized by means with SDs, median with ranges, or frequencies with percentages. Summary statistics were compared between groups using ANOVA, independent 1-tailed *t* tests, median regression, Wilcoxon rank sum, Fisher exact, or chi-square tests according to the type and distribution of each variable of interest. Pairwise tests with Holm *P* value adjustment were done when necessary. Since patients had multiple visits, some comparisons were performed using mixed effects linear regression or generalized estimating equations (GEEs) to account for the correlated data structure as appropriate. GEE models used an exchangeable correlation structure, when possible; otherwise, an independent correlation structure was used. A multiple GEE model was used to determine whether not showing up was associated with visit-type appointments. We included covariates that had statistical significance associated with a no-show or visit type (telemedicine vs in-person). For example, older age and certain medications (methotrexate and belimumab) were more associated with telemedicine visits and introduced into the GEE model. Among patients seen both in 2019 and 2020, GEE was used to determine factors associated with laboratory test adherence (defined as being performed within 1 month of the visit). For laboratory test visits, the first visit was used if a patient had multiple visits less than 6 weeks apart. The GEE models estimated odds ratios (ORs) with 95% CI. We performed the analysis of “no-shows” on our subgroup of patients that were seen in both 2019 and 2020 to account for the same patients that historically followed.

### Ethical Considerations

This retrospective study was approved by the Baylor College of Medicine Institutional Review Board with waived informed consent under protocol number H-45296. As this was a retrospective review, a waiver of consent was granted. Our data were deidentified and all results were stored in secure and encrypted servers at the Baylor College of Medicine. The data collected were deidentified prior to analysis.

## Results

### Baseline Characteristics

Baseline characteristics of included patients are shown in Table 1.

There were 156 patients included in our analysis. Most patients were female (90.4%), were Hispanic (49.3%), had a median age of 43 (range 19-80) years, and had received hydroxychloroquine (n=144, 92%) or prednisone (n=120, 77%) throughout the follow-up period. Baseline characteristics broken down by visit type are shown in Table S1 in Multimedia Appendix 1. We included 771 in-person visits and 182 telemedicine (including telephone or video) visits in our analysis. We found that telemedicine visits were associated with older age (median 45.3, range 19.9-81.5 vs 41.2, range 19.2-82.1 years; *P*=.01) and were less likely to occur in patients who were prescribed mycophenolate (74/182, 40.7% vs 367/771, 47.6%; *P*=.03) or prednisone (142/182, 78% vs 644/771, 83.5%; *P*=.02) compared to in-person visits. Differences in all other characteristics were not statistically significant (all *P*>.05; Table S2 in Multimedia Appendix 1).

**Table 1.** Baseline characteristics of patients with SLE (N=156).

| Baseline characteristics | Values |
|---|---|
| **Sex, n (%)** | |
| Female | 141 (90.4) |
| Male | 15 (9.6) |
| **Race (n=150), n (%)** | |
| Hispanic | 74 (49.3) |
| Non-Hispanic Black | 35 (23.3) |
| Non-Hispanic White | 17 (11.3) |
| Other | 24 (16) |
| Age (years), median (range) | 43.2 (19.2-79.5) |
| **Prescription drug use (ever prescribed as per medical record), n (%)** | |
| Hydroxychloroquine | 144 (92.3) |
| Mycophenolate | 63 (40.4) |
| Azathioprine | 41 (26.3) |
| Methotrexate | 23 (14.7) |
| Rituximab | 4 (2.6) |
| Belimumab | 16 (10.3) |
| Tacrolimus | 8 (5.1) |
| Prednisone | 120 (77) |
| Cyclophosphamide | 5 (3.2) |

## Patient Characteristics Associated With Visit Type

We determined differences in patient characteristics between those who had all in-person visits, 2 or more telemedicine visits, or an in-person visit and 1 telemedicine visit in 2020 (Table 2).

Patients who had 2 or more telemedicine appointments were less likely to be prescribed methotrexate (3/63, 4.8% vs 11/54, 20.4%; Holm adjusted *P* value=.03) or prednisone (27/63, 42.9% vs 37/54, 68.5%; Holm adjusted *P* value=.03) during the year compared to those with in-person appointments or only 1 telemedicine appointment. Whether the patient was prescribed mycophenolate was significantly different between visit types (*P*=.05), but when performing pairwise comparisons on each visit type category, none of them were significant (all *P*>.05). No patients had received cyclophosphamide in this study.

**Table 2.** Patient characteristics associated with visit types in 2020.

| Patient characteristics | All in-person visits (n=21) | 2 or more telemedicine visits[a] (n=63) | In-person visit + 1 telemedicine visit (n=54) | *P* value[b] | *P* value , significant pairwise comparison results[c] |
|---|---|---|---|---|---|
| **Age (years; first 2020 visit), median (IQR)** | 36.7 (33.1-53) | 45.1 (20.6-77.0) | 45.2 (19.8-81.5) | .25 | N/A[d] |
| **Sex, n (%)** | | | | | |
| Female | 19 (90.5) | 59 (93.7) | 49 (90.7) | .76 | N/A |
| **Race, n (%)[e]** | | | | .08 | |
| Hispanic | 6 (30) | 32 (52.5) | 25 (49) | | N/A |
| Non-Hispanic Black | 9 (45) | 13 (21.3) | 11 (21.6) | | N/A |
| Non-Hispanic White | 4 (20) | 3 (4.9) | 5 (9.8) | | N/A |
| Others | 1 (5) | 13 (21.3) | 10 (19.6) | | N/A |
| **Medications, n (%)** | | | | | |
| Hydroxychloroquine | N/A | 47 (74.6) | 44 (81.5) | .67 | N/A |
| Mycophenolate | 9 (42.9) | 15 (23.8) | 24 (44.4) | .05 | NS[f] |
| Azathioprine | 5 (23.8) | 11 (17.5) | 14 (25.9) | .52 | N/A |
| Methotrexate | 3 (14.3) | 3 (4.8) | 11 (20.4) | .02 | .03[g] |
| Rituximab | 0 (0) | 1 (1.6) | 1 (1.9) | >.99 | N/A |
| Belimumab | 1 (4.8) | 3 (4.8) | 2 (3.7) | >.99 | N/A |
| Tacrolimus | 0 (0.0) | 3 (4.8) | 2 (3.7) | .85 | N/A |
| Prednisone | 15 (71.4) | 27 (42.9) | 37 (68.5) | .008 | .03[g] |

[a]This includes patients who have only 1 visit in 2020, and that visit is a telemedicine visit.

[b]Calculated using median regression test, Fisher exact test, or chi-square test.

[c]Pairwise Fisher exact test with Holm *P* value adjustment.

[d]N/A: not applicable.

[e]All in-person visits: n=20; two or more telemedicine visits: n=61; and in-person visit + 1 telemedicine visit: n=51.

[f]NS: Not significant.

[g]Two or more telemedicine visits versus in-person visit + 1 telemedicine visit.

## No-Shows for Clinic Visits

All clinic visits from March to September in 2018 and 2019 were in-person (275 in 2018 and 305 in 2019). From March 2020 to September 2020, of out a total of 373 visits, there were 191 (51.2%) in-person visits and 182 (48.8%) telemedicine visits (Table 3). There was no statistical difference in the no-show rates between in-person visits in 2018, 2019, and 2020 (31/275, 11.3% vs 38/305, 12.5% vs 31/191, 16.2%, respectively). In 2020, when telemedicine was implemented, the no-show rate for in-person visits was 16.2% (31/191) versus 5.5% (10/182) for telemedicine visits (*P*=.002). We used independent GEEs to determine any characteristics associated with no-shows (Table S3 in Multimedia Appendix 1). After adjusting for age and significant SLE prescription drugs (methotrexate and belimumab) in a multiple GEE, there was a significantly decreased odds of no-shows for telemedicine versus in-person clinic appointments (adjusted OR 0.39, 95% CI 0.20-0.77; *P*=.007; Table S4 in Multimedia Appendix 1).

We also performed a subgroup analysis on patients who were seen at least once in 2019 and 2020. There were 300 visits in 2019 and 332 visits in 2020. The total no-show rates between 2019 and 2020 were similar (38/300, 11% vs 41/332, 10.5%; *P*=.85). Among these visits, we also found that telemedicine appointments had significantly lower odds of no-show compared to in-person appointments (adjusted OR 0.31, 95% CI 0.14-0.69) when adjusting for those SLE prescription drugs that were significantly different (only rituximab) according to the type of visit.

**Table 3.** Visit characteristics stratified by type of visit.

| Visit characteristics | In-person visits (n=771) | Telephone visits (n=157) | Video visits (n=25) | Telephone or video visits (n=182) | Total, n (%) |
|---|---|---|---|---|---|
| **Year, n (%)** | | | | | |
| 2018 (n=275) | 275 (100) | 0 (0) | 0 (0) | 0 (0) | 275 (100) |
| 2019 (n=305) | 305 (100) | 0 (0) | 0 (0) | 0 (0) | 305 (100) |
| 2020 (n=373) | 191 (51.2) | 157 (42.1) | 25 (6.7) | 182 (48.8) | 373 (100) |
| **No-shows, n/N (%)** | | | | | |
| 2018 | 31/275 (11.3) | N/A[a] | N/A | N/A | 31/275 (11.3) |
| 2019 | 38/305 (12.5) | N/A | N/A | N/A | 38/305 (12.5) |
| 2020 | 31/191 (16.2) | 7/157 (4.5) | 3/25 (12) | 10/182 (5.5) | 41/373 (11) |

[a]N/A: not applicable.

### Laboratory Test Use

When comparing laboratory test use between 2019 and 2020, the only significant difference was in urinalysis which was more frequently performed for telemedicine visits than in-person visits (38/289, 13.1% vs 7/257, 2.7%; *P*<.001; Table 4). We also compared the use of laboratory tests between in-person and telemedicine visits in 2020 using GEE. No statistically significant differences were observed (all *P*>.05). We found that there were no differences in nonadherence to laboratory testing for all laboratory tests, although there was a trend toward significance for anti-dsDNA testing (4/136, 2.9% nonadherence for in-person visits vs 13/153, 8.5% nonadherence for telemedicine visits; *P*=.06; Table 4), but it was not statistically significant. We found that urine studies had the highest proportion of nonadherence (16/136, 11.8% for in-person visits vs 22/153, 14.4% for telemedicine visits; *P*=.51), although this could be explained by other factors not measured such as end-stage renal disease.

**Table 4.** Nonadherence to laboratory testing for completed visits in 2019 and 2020.

| Laboratory studies not completed within 30 days of appointment | Total visits in 2019 (n=257), n (%) | Total visits in 2020 (n=289), n (%) | *P* value[a] | In-person visits in 2020 (n=136), n (%) | Telemedicine visits in 2020 (n=153), n (%) | *P* value[a] |
|---|---|---|---|---|---|---|
| CBC[b] | 7 (2.7) | 4 (1.4) | .28 | 0 (0) | 4 (2.6) | N/A[c] |
| BMP[d] or CMP[e] | 5 (1.9) | 7 (2.4) | .71 | 2 (1.5) | 5 (5.3) | .33 |
| Urinalysis | 7 (2.7) | 38 (13.1) | <.001 | 16 (11.8) | 22 (14.4) | .51 |
| Anti-dsDNA[f] | 19 (7.4) | 17 (5.9) | .48 | 4 (2.9) | 13 (8.5) | .06 |
| Complements | 9 (3.5) | 18 (6.2) | .15 | 5 (3.7) | 13 (8.5) | .10 |

[a]Calculated using a generalized estimating equation with an independent correlation structure.

[b]CBC: complete blood count.

[c]N/A: not applicable.

[d]BMP: basic metabolic panel.

[e]CMP: comprehensive metabolic panel.

[f]Anti-dsDNA: anti–double-stranded DNA.

## Discussion

We evaluated adherence to telemedicine visits in the management of patients with SLE, at a publicly funded county hospital serving primarily underserved patients. We also determined whether there were differences in laboratory use between patients who received telemedicine versus in-person visits. Our results demonstrate that telemedicine encounters had significantly lower odds of no-shows compared to in-person encounters. We also found that no-show rates were similar for 2019 and 2020 despite the emergence of the COVID-19 pandemic, which could be due to the availability of telemedicine, as no-shows for telemedicine versus in-person in 2020 were significantly lower (10/182, 5.5% vs 31/191, 16.2%). Furthermore, to our knowledge, this is the first study that shows telemedicine visits do not affect laboratory use within 30 days of the clinic visits.

Studies have shown that telemedicine can play a role in the management of chronic diseases that require frequent clinic visits [14]. Other studies in SLE have shown that telemedicine was used as frequently as in-person visits during the initial COVID-19 pandemic, although this is the first study to demonstrate that this occurred in an underresourced patient population [15,16]. Of note, the widespread use of telemedicine is also seen in patients with severe chronic diseases such as SLE. A recently published randomized controlled trial in Hong

Kong found that the use of telemedicine in patients with lupus nephritis was associated with more hospitalizations [17]. Our study did not address disease activity or hospitalization, and further research is needed to assess how the widespread use of telemedicine may impact these factors.

Our study is consistent with studies in other populations that suggest that telemedicine may provide advantages for underserved populations by decreasing missed appointments. One systematic review of 28 studies reported on the use of telehealth for patients from racial and ethnic minority populations. Results showed that the implementation of telehealth improved access to care; however, there were still barriers related to the technology needed for telemedicine [13]. A separate study using administrative claims data also examined the use of telemedicine in general patient populations and found that telemedicine was associated with fewer missed appointments [18]. However, this study did not include patient populations such as SLE that require frequent clinic visits and laboratory studies (at least every 3 to 4 months). Although our study suggests that telemedicine may be a strategy to decrease the no-show rate in patients with low SES and SLE, more research is needed to determine how other characteristics (including primary language, digital literacy, and disease activity) influence telemedicine and potentially disease outcomes. Furthermore, telemedicine should now be studied as the COVID-19 pandemic has entered the endemic phase.

The strength of our study includes a large number of patients with SLE of low SES in 1 large hospital system where all clinic appointments and laboratory values are documented. There are several limitations in our study. First, we had a predominance of telephone encounters compared to video visits, albeit this has also been seen in other studies, especially among patients of Black and Hispanic ethnicity and of low SES, which were the majority [19]. The use of video visits may affect the no-show rate by presenting technological challenges. Second, our study did not adjust for disease severity according to validated indices as it was retrospective, and it only used prescription drug use as a surrogate for severity. Finally, we were unable to adequately control for insurance type at the time of the scheduled appointment as this information is not updated regularly; however, we do not expect this to change the results as over 85% of patients in the HHS are publicly insured or uninsured.

In conclusion, our study shows that the use of telemedicine during the initial phase of the COVID-19 pandemic was associated with a low rate of no-shows in a population of underserved patients with SLE without impacting laboratory use. To our knowledge, this is the first study to demonstrate that in patients with SLE telemedicine is not associated with decreased laboratory screening, which is a critical component to the care of patients with lupus. As such, we believe the results of this study warrant further investigation to determine the clinical impact of telemedicine on SLE in prospective studies, as the design of this study was not able to capture important clinical characteristics that may influence telemedicine and clinical outcomes, including digital literacy and disease activity.

## Acknowledgments

No artificial intelligence was used in this study or manuscript.

## Data Availability

The datasets generated during this study are not publicly available because the participants of this study did not give written consent for their data to be shared publicly but are available from the corresponding author on reasonable request.

## Authors' Contributions

SB, MESA, and SKA conceptualized the study and study protocol. SB collected the data, and KAS did all the statistical analysis. Interpretation of data was done by all the authors. The manuscript was written by SB and reviewed by all authors. All authors reviewed the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics by visit type, *P* value testing results, and multiple generalized estimating equation odds ratios for no-shows.
[DOCX File , 24 KB - ijmr_v13i1e49065_app1.docx ]

## References

1. Keesara S, Jonas A, Schulman K. COVID-19 and health care's digital revolution. N Engl J Med 2020;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
2. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly B, et al. Trends in the use of telehealth during the emergence of the COVID-19 pandemic—United States, January-March 2020. MMWR Morb Mortal Wkly Rep 2020;69(43):1595-1599 [FREE Full text] [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]

3. Ugarte-Gil M, Alarcón GS, Izadi Z, Duarte-García A, Reátegui-Sokolova C, Clarke AE, et al. Characteristics associated with poor COVID-19 outcomes in individuals with systemic lupus erythematosus: data from the COVID-19 global rheumatology alliance. Ann Rheum Dis 2022;81(7):970-978 [FREE Full text] [doi: 10.1136/annrheumdis-2021-221636] [Medline: 35172961]
4. Raiker R, Pakhchanian H, DeYoung C, Gupta L, Kardeş S, Ahmed S, et al. Short term outcomes of COVID-19 in lupus: propensity score matched analysis from a nationwide multi-centric research network. J Autoimmun 2021;125:102730 [FREE Full text] [doi: 10.1016/j.jaut.2021.102730] [Medline: 34649723]
5. McQueenie R, Ellis DA, McConnachie A, Wilson P, Williamson AE. Morbidity, mortality and missed appointments in healthcare: a national retrospective data linkage study. BMC Med 2019;17(1):2 [FREE Full text] [doi: 10.1186/s12916-018-1234-0] [Medline: 30630493]
6. Petri M, Perez-Gutthann S, Longenecker J, Hochberg M. Morbidity of systemic lupus erythematosus: role of race and socioeconomic status. Am J Med 1991;91(4):345-353 [FREE Full text] [doi: 10.1016/0002-9343(91)90151-m] [Medline: 1951378]
7. Li D, Madhoun HM, Roberts WN, Jarjour W. Determining risk factors that increase hospitalizations in patients with systemic lupus erythematosus. Lupus 2018;27(8):1321-1328. [doi: 10.1177/0961203318770534] [Medline: 29667857]
8. Darrat I, Tam S, Boulis M, Williams AM. Socioeconomic disparities in patient use of telehealth during the coronavirus disease 2019 surge. JAMA Otolaryngol Head Neck Surg 2021;147(3):287-295 [FREE Full text] [doi: 10.1001/jamaoto.2020.5161] [Medline: 33443539]
9. Jaffe DH, Lee L, Huynh S, Haskell TP. Health inequalities in the use of telehealth in the United States in the lens of COVID-19. Popul Health Manag 2020;23(5):368-377. [doi: 10.1089/pop.2020.0186] [Medline: 32816644]
10. Gladman DD, Ibañez D, Ruiz I, Urowitz MB. Recommendations for frequency of visits to monitor systemic lupus erythematosus in asymptomatic patients: data from an observational cohort study. J Rheumatol 2013;40(5):630-633. [doi: 10.3899/jrheum.121094] [Medline: 23457385]
11. Quick facts: Harris County, Texas. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/harriscountytexas/POP010220#POP010220.) [accessed 2024-08-22]
12. Duarte-García A, Fang H, To CH, Magder LS, Petri M. Seasonal variation in the activity of systemic lupus erythematosus. J Rheumatol 2012;39(7):1392-1398 [FREE Full text] [doi: 10.3899/jrheum.111196] [Medline: 22660806]
13. Truong M, Yeganeh L, Cook O, Crawford K, Wong P, Allen J. Using telehealth consultations for healthcare provision to patients from non-indigenous racial/ethnic minorities: a systematic review. J Am Med Inform Assoc 2022;29(5):970-982 [FREE Full text] [doi: 10.1093/jamia/ocac015] [Medline: 35150266]
14. Chen MH, Goverover Y, Botticello A, DeLuca J, Genova HM. Healthcare disruptions and use of telehealth services among people with multiple sclerosis during the COVID-19 pandemic. Arch Phys Med Rehabil 2022;103(7):1379-1386 [FREE Full text] [doi: 10.1016/j.apmr.2021.12.028] [Medline: 35093328]
15. Kasturi S, Price LL, Paushkin V, Salmon JE, McAlindon TE, Mandl LA. Impact of the first wave of the COVID-19 pandemic on systemic lupus erythematosus patients: results from a multi-center prospective cohort. Lupus 2021;30(11):1747-1755 [FREE Full text] [doi: 10.1177/09612033211033981] [Medline: 34284676]
16. Cavagna L, Zanframundo G, Codullo V, Pisu M, Caporali R, Montecucco C. Telemedicine in rheumatology: a reliable approach beyond the pandemic. Rheumatology 2021;60(1):366-370 [FREE Full text] [doi: 10.1093/rheumatology/keaa554] [Medline: 32893293]
17. So H, Chow E, Cheng IT, Lau S, Li TK, Szeto C, et al. Use of telemedicine for follow-up of lupus nephritis in the COVID-19 outbreak: the 6-month results of a randomized controlled trial. Lupus 2022;31(4):488-494 [FREE Full text] [doi: 10.1177/09612033221084515] [Medline: 35254169]
18. Adepoju OE, Chae M, Liaw W, Angelocci T, Millard P, Matuk-Villazon O. Transition to telemedicine and its impact on missed appointments in community-based clinics. Ann Med 2022;54(1):98-107 [FREE Full text] [doi: 10.1080/07853890.2021.2019826] [Medline: 34969330]
19. Huang J, Graetz I, Millman A, Gopalan A, Lee C, Muelly E, et al. Primary care telemedicine during the COVID-19 pandemic: patient's choice of video versus telephone visit. JAMIA Open 2022;5(1):ooac002 [FREE Full text] [doi: 10.1093/jamiaopen/ooac002] [Medline: 35146380]


## Abbreviations

**Anti-dsDNA:** anti–double-stranded DNA
**GEE:** generalized estimating equation
**HHS:** Harris Health System
**OR:** odds ratio
**SES:** socioeconomic status
**SLE:** systemic lupus erythematosus

*Edited by T de Azevedo Cardoso; submitted 16.05.23; peer-reviewed by SR Patlolla; comments to author 06.02.24; revised version received 13.02.24; accepted 20.05.24; published 01.11.24.*

*Please cite as:*
*Bruera S, Staggers KA, Suarez-Almazor ME, Agarwal SK*
*Telemedicine for Patients With Systemic Lupus Erythematosus in a Publicly Funded Hospital System: Retrospective Study*
*Interact J Med Res 2024;13:e49065*
*URL: https://www.i-jmr.org/2024/1/e49065*
*doi:10.2196/49065*
*PMID:39078399*

Original Paper

# Behavioral Insights from Vaccine Adoption in Nigeria: Cross-Sectional Survey Findings

Sohail Agha[1,2], BA, PhD; Ifeanyi Nsofor[3], MSc, MD; Drew Bernard[2], BA; Sarah Francis[2], BA; Nandan Rao[4], MSc

[1]Behavior Design Lab, Stanford University, Stanford, CA, United States

[2]Behavioral Insights Lab, Seattle, WA, United States

[3]Centre for Family Health Initiative, Abuja, Nigeria

[4]Virtual 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.

***(Interact J Med Res 2024;13:e47817)*** doi:10.2196/47817

**KEYWORDS**

behavioral insights; COVID-19; Nigeria; surveys; vaccination

## 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 vaccination 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.

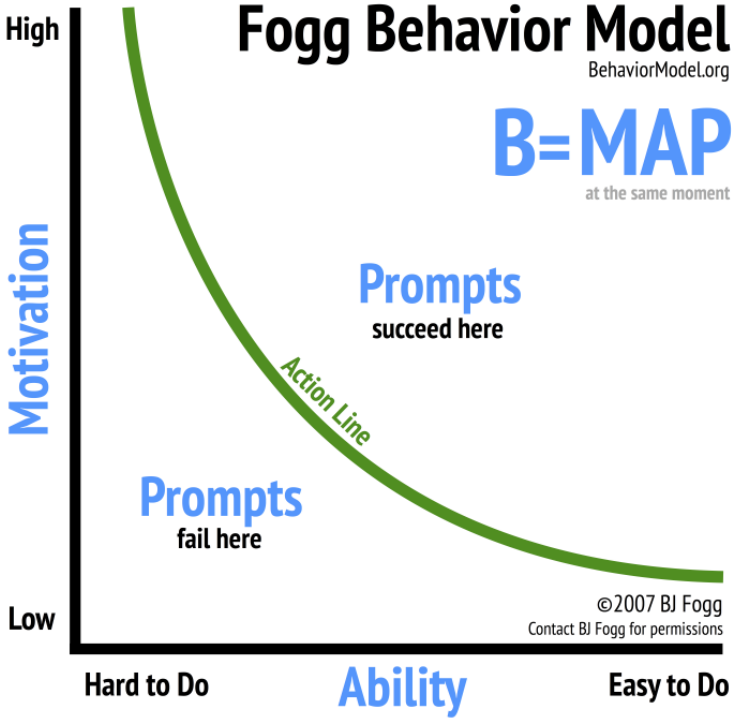
The practitioner-friendly model used in this study, the Fogg Behavior Model (FBM), was introduced in the public health literature in 2019 to explain the effects of a social marketing behavior change campaign on the adoption of condoms by married men in Pakistan [13]. More recently, it has been used to 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].

**Figure 1.** Fogg Behavior Model (reproduced from Fogg [22], with permission from BJ Fogg).




## 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

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.

| Core motivators | Frequency distributions of core motivators (N=957, 100%), n (%) | Nigerian adults who obtained at least 1 vaccination (n=540, 56.4%), n (%) | *P* value | Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)[a] |
|---|---|---|---|---|
| **Hope: Vaccination allows a person to live a better life** | | | <.001 | |
| Agree or strongly agree | 726 (75.9) | 483 (66.5) | | 3.40 (2.27-5.09) |
| Disagree or strongly disagree or do not know | 231 (24.1) | 57 (24.7) | | 1.00 (Reference) |
| **Fear: Getting vaccinated protects people from hardship** | | | <.001 | |
| Agree or strongly agree | 444 (46.4) | 301 (67.8) | | 1.09 (0.78-1.51) |
| Disagree or strongly disagree or do not know | 513 (53.6) | 239 (46.6) | | 1.00 (Reference) |
| **Pleasure: Getting vaccinated allows people to do more things they enjoy** | | | <.001 | |
| Agree or strongly agree | 610 (63.7) | 411 (67.4) | | 1.84 (1.30-2.62) |
| Disagree or strongly disagree or do not know | 347 (36.3) | 129 (37.2) | | |
| **Pain: Getting vaccinated reduces the likelihood of getting or spreading COVID-19** | | | <.001 | |
| Agree or strongly agree | 780 (81.5) | 477 (61.2) | | 1.15 (0.76-1.73) |
| Disagree or strongly disagree or do not know | 177 (18.5) | 63 (35.6) | | 1.00 (Reference) |
| **Acceptance: Many friends and family approve vaccination** | | | <.001 | |
| Agree or strongly agree | 757 (79.1) | 472 (62.4) | | 1.76 (1.21-2.58) |
| Disagree or strongly disagree or do not know | 200 (20.9) | 68 (34) | | 1.00 (Reference) |
| **Rejection: Most family and friends do not approve of vaccination** | | | .56 | |
| Agree or strongly agree | 355 (37.1) | 344 (57.1) | | 1.06 (0.79-1.43) |
| Disagree or strongly disagree or do not know | 602 (62.9) | 196 (55.2) | | 1.00 (Reference) |

[a]Pseudo $R^2$=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). Respondents' concern that not 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

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=382) of respondents felt that the physical effort required made it difficult to get vaccinated. Nearly half (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.

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

[a]Pseudo $R^2$=12%.

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.

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

[a]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.

| | Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)[a] |
|---|---|
| **Core motivators** | |
| Getting vaccinated allows a person to live a better life (hope) | 3.51 (2.23-5.52) |
| Getting vaccinated protects people from hardship (fear) | 1.03 (0.71-1.49) |
| Allows people to do more things they enjoy (pleasure) | 1.97 (1.33-2.93) |
| Reduces the likelihood of getting or spreading COVID-19 (pain) | 1.17 (0.74-1.84) |
| Many friends and family approve vaccination (acceptance) | 1.62 (1.06-2.48) |
| Most family and friends do not approve vaccination (rejection) | 1.18 (0.83-1.67) |
| **Ability factors** | |
| Family or work responsibilities make it difficult (time) | 0.85 (0.61-1.19) |
| Costs or loss of income make it difficult (money) | 0.35 (0.24-0.50) |
| Physical effort makes it difficult (physical effort) | 1.45 (1.02-2.07) |
| The decision to get the vaccine is difficult (mental effort) | 1.08 (0.77-1.51) |
| Vaccine not available where I routinely visit (routine) | 0.29 (0.21-0.40) |
| **Sociodemographic factors** | |
| **Age (years)** | |
| 18-29 | 1.09 (0.68-1.74) |
| 30-39 | 1.14 (0.70-1.86) |
| ≥40 | 1.00 (Reference) |
| **Sex** | |
| Male | 1.00 (Reference) |
| Female | 0.80 (0.58-1.10) |
| **Education** | |
| Primary or secondary school certificate | 1.00 (Reference) |
| Ordinary national diploma (OND) | 2.27 (1.34-3.84) |
| Higher national diploma (HND) | 3.57 (2.04-6.24) |
| Bachelors or higher | 1.26 (0.86-1.85) |
| Other | 1.88 (0.82-4.33) |
| **Location** | |
| City | 1.00 (Reference) |
| Town | 0.90 (0.64-1.26) |
| Rural | 0.54 (0.31-0.92) |

[a]Pseudo $R^{2=}$23.51%.

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

1. Robinson ED, Wilson P, ELeki BJ, Wonodi W. Knowledge, acceptance, and hesitancy of COVID-19 vaccine among health care workers in Nigeria. MGM J Med Sci 2021;8(2):102-110 [FREE Full text] [doi: 10.4103/mgmj.mgmj_4_21]
2. Osuagwu UL, Mashige KP, Ovenseri-Ogbomo G, Envuladu EA, Abu EK, Miner CA, et al. The impact of information sources on COVID-19 vaccine hesitancy and resistance in sub-Saharan Africa. BMC Public Health 2023;23(1):38 [FREE Full text] [doi: 10.1186/s12889-022-14972-2] [Medline: 36609264]
3. Adejumo OA, Ogundele OA, Madubuko CR, Oluwafemi RO, Okoye OC, Okonkwo KC, et al. Perceptions of the COVID-19 vaccine and willingness to receive vaccination among health workers in Nigeria. Osong Public Health Res Perspect 2021;12(4):236-243 [FREE Full text] [doi: 10.24171/j.phrp.2021.0023] [Medline: 34289295]
4. Iliyasu Z, Umar AA, Abdullahi HM, Kwaku AA, Amole TG, Tsiga-Ahmed FI, et al. "They have produced a vaccine, but we doubt if COVID-19 exists": correlates of COVID-19 vaccine acceptability among adults in Kano, Nigeria. Hum Vaccin Immunother 2021;17(11):4057-4064 [FREE Full text] [doi: 10.1080/21645515.2021.1974796] [Medline: 34613864]
5. Babafunke F, Olajumoke O, Oluwatofunmi L, Oluyemi A, Oluseyi A, Oluwagbemisola F, et al. Perceived risk and protective behaviors related to COVID-19 among the Nigerian general population: an application of the extended parallel process model. Sci Afr 2022;16:e01259 [FREE Full text] [doi: 10.1016/j.sciaf.2022.e01259] [Medline: 35765588]
6. Allagoa DO, Oriji PC, Tekenah ES, Obagah L, Njoku C, Afolabi AS, et al. Predictors of acceptance of COVID-19 vaccine among patients at a tertiary hospital in South-South Nigeria. Int J Community Med Public Health 2021;8(5):2165-2172 [FREE Full text] [doi: 10.18203/2394-6040.ijcmph20211733]
7. Ilesanmi OS, Akosile PO, Afolabi AA, Ukwenya VO. Handling distrust on risk communication in Nigeria: a strategy to strengthening the COVID-19 outbreak response. Disaster Med Public Health Prep 2021;17:e27 [FREE Full text] [doi: 10.1017/dmp.2021.245] [Medline: 34308825]
8. Babatope T, Ilyenkova V, Marais D. COVID-19 vaccine hesitancy: a systematic review of barriers to the uptake of COVID-19 vaccine among adults in Nigeria. Bull Natl Res Cent 2023;47(1):45 [FREE Full text] [doi: 10.1186/s42269-023-01017-w] [Medline: 36970323]
9. Lawal O. Understanding risk communication effectiveness from public interest, mobility, and COVID-19 cases: a case study of COVID-19 in Nigeria. Front Commun 2022;7:921648 [FREE Full text] [doi: 10.3389/fcomm.2022.921648]
10. Olu-Abiodun O, Abiodun O, Okafor N. COVID-19 vaccination in Nigeria: a rapid review of vaccine acceptance rate and the associated factors. PLoS One 2022;17(5):e0267691 [FREE Full text] [doi: 10.1371/journal.pone.0267691] [Medline: 35544545]
11. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
12. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
13. Agha S, Tollefson D, Paul S, Green D, Babigumira JB. Use of the fogg behavior model to assess the impact of a social marketing campaign on condom use in Pakistan. J Health Commun 2019;24(3):284-292. [doi: 10.1080/10810730.2019.1597952] [Medline: 30945612]
14. Evans WD, Gerard R, Symington L, Shaikh H, Agha S. Implementation practice models for development in low- and middle-income countries: systematic review of peer-reviewed literature. BMC Public Health 2022;22(1):1157 [FREE Full text] [doi: 10.1186/s12889-022-13530-0] [Medline: 35681165]
15. Agha S. Use of a practitioner-friendly behavior model to identify factors associated with COVID-19 vaccination and other behaviors. Vaccines (Basel) 2022;10(8):1261 [FREE Full text] [doi: 10.3390/vaccines10081261] [Medline: 36016149]
16. Pedersen B, Thanel K, Kouakou AY, Zo JR, Ouattara ML, Gbeke D, et al. Identifying drivers of COVID-19 vaccine uptake among residents of Yopougon Est, Abidjan, Côte d'Ivoire. Vaccines (Basel) 2022;10(12):2101 [FREE Full text] [doi: 10.3390/vaccines10122101] [Medline: 36560511]
17. Agha S, Chine A, Lalika M, Pandey S, Seth A, Wiyeh A, et al. Drivers of COVID-19 vaccine uptake amongst healthcare workers (HCWs) in Nigeria. Vaccines (Basel) 2021;9(10):1162 [FREE Full text] [doi: 10.3390/vaccines9101162] [Medline: 34696270]
18. Evans WD, Bingenheimer JB, Long M, Ndiaye K, Donati D, Rao NM, et al. Development and evaluation of an index to measure the ability to get vaccinated for COVID-19. Vaccines (Basel) 2023;11(2):342 [FREE Full text] [doi: 10.3390/vaccines11020342] [Medline: 36851220]
19. Agha S, Morgan B, Archer H, Paul S, Babigumira JB, Guthrie BL. Understanding how social norms affect modern contraceptive use. BMC Public Health 2021;21(1):1061 [FREE Full text] [doi: 10.1186/s12889-021-11110-2] [Medline: 34088295]
20. Meekers D, Onuoha C, Olutola O. Applying the Fogg Behavior Model to improve contraceptive social marketing during the COVID-19 lockdown in Nigeria: a case study. Gates Open Res 2020;4:141 [FREE Full text] [doi: 10.12688/gatesopenres.13186.1]
21. Fogg B. A behavior model for persuasive design. New York, NY, US: Association for Computing Machinery; 2009 Presented at: Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, California, USA p. 1-7. [doi: 10.1145/1541948.1541999]

22. Fogg BJ. Tiny Habits: The Small Changes that Change Everything. Boston, MA: Houghton Mifflin Harcourt; 2020.
23. Rao N, Donati D, Orozco V. Conducting Surveys and Interventions Entirely Online: A Virtual Lab Practitioner's Manual. The World Bank. 2020. URL: https://documents1.worldbank.org/curated/en/819191635501106862/pdf/Conducting-Surveys-and-Interventions-Entirely-Online-A-Virtual-Lab-Practitioner-s-Manual.pdf [accessed 2024-01-23]
24. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
25. Shaver LG, Khawer A, Yi Y, Aubrey-Bassler K, Etchegary H, Roebothan B, et al. Using Facebook advertising to recruit representative samples: feasibility assessment of a cross-sectional survey. J Med Internet Res 2019;21(8):e14021 [FREE Full text] [doi: 10.2196/14021] [Medline: 31429409]
26. Pedersen B, Thompson G, Kouakou AY, Mujinga M, Nicholes S, Martinez A, et al. Development and assessment of a six-item index to gauge motivation to receive COVID-19 vaccination. Vaccines 2023;12(1):6 [FREE Full text] [doi: 10.3390/vaccines12010006]

## 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

*Edited by T Leung; submitted 02.04.23; peer-reviewed by J Smith, OO Oyewole; comments to author 27.04.23; revised version received 24.05.23; accepted 02.01.24; published 26.02.24.*

*Please cite as:*
*Agha S, Nsofor I, Bernard D, Francis S, Rao N*
*Behavioral Insights from Vaccine Adoption in Nigeria: Cross-Sectional Survey Findings*
*Interact J Med Res 2024;13:e47817*
*URL: https://www.i-jmr.org/2024/1/e47817*
*doi:10.2196/47817*
*PMID:38407956*

Original Paper

# Influence of Environmental Factors and Genome Diversity on Cumulative COVID-19 Cases in the Highland Region of China: Comparative Correlational Study

Zhuoga Deji[1,2], MSc; Yuantao Tong[1], PhD; Honglian Huang[1,3], PhD; Zeyu Zhang[1], PhD; Meng Fang[4], PhD; M James C Crabbe[5,6,7], PhD; Xiaoyan Zhang[1], PhD; Ying Wang[3], PhD

[1]Research Center for Translational Medicine, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China

[2]Information School, The University of Sheffield, Sheffield, United Kingdom

[3]Department of Clinical Laboratory Medicine Center, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[4]Department of Laboratory Medicine, Shanghai Eastern Hepatobiliary Surgery Hospital, Shanghai, China

[5]Wolfson College, Oxford University, Oxford, United Kingdom

[6]Institute of Biomedical and Environmental Science & Technology, University of Bedfordshire, Bedfordshire, United Kingdom

[7]School 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.

***(Interact J Med Res 2024;13:e43585)*** doi:10.2196/43585

**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:

$$\text{Population density} = \text{population size/area of the land } (\text{km}^2) \quad \textbf{(1)}$$

The normalized daily number of confirmed COVID-19 cases was calculated using the following equation:

$$\text{Normalized COVID-19 cases} = \text{total number of confirmed cases/population density} \quad \textbf{(2)}$$

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 Kolmogorow-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:

$$Y = X\beta + Z\mu + \in \quad \textbf{(3)}$$

where Xβ represents the fixed-effects set in this study, including the covariates meteorological and air quality, and Zμ represents an N×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 2.17-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 (Sn) 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 Sn according to the following formula [25]:

$$Sn = i\sum(pilnpi)/lnN \quad \textbf{(4)}$$

where pi 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” [26] (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” [27] (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.

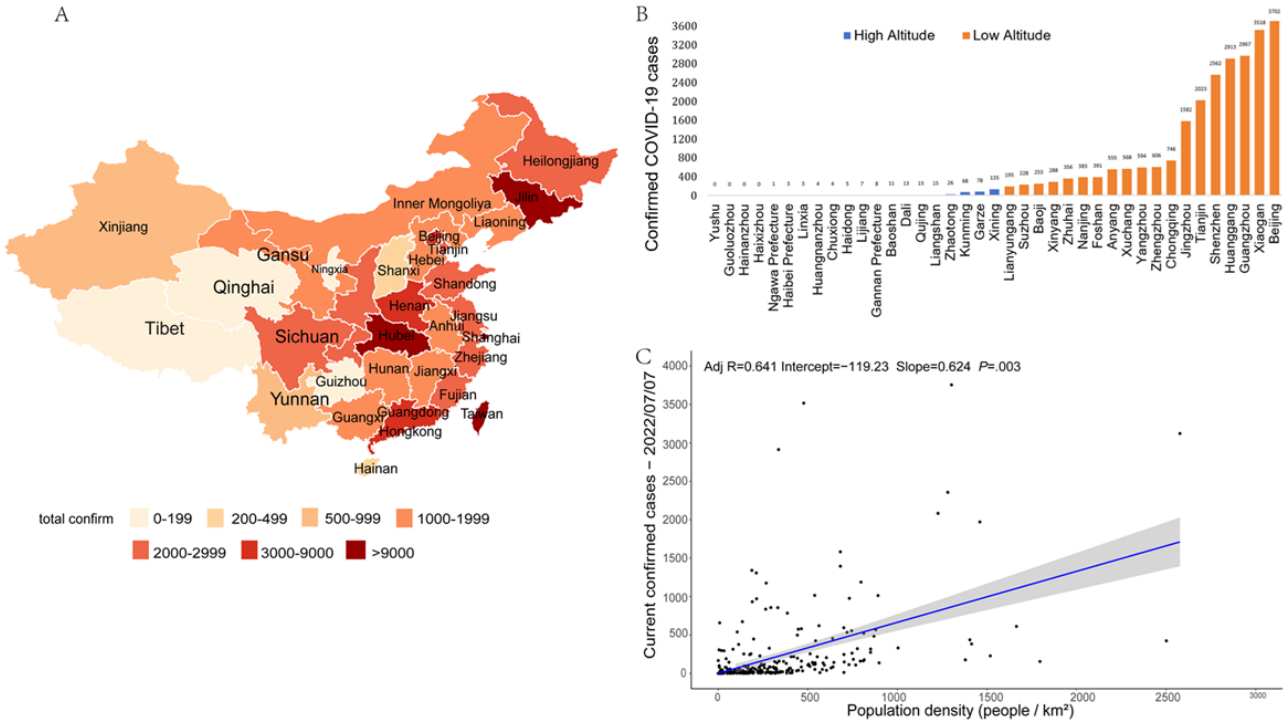
## Results

### 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

1B). 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.

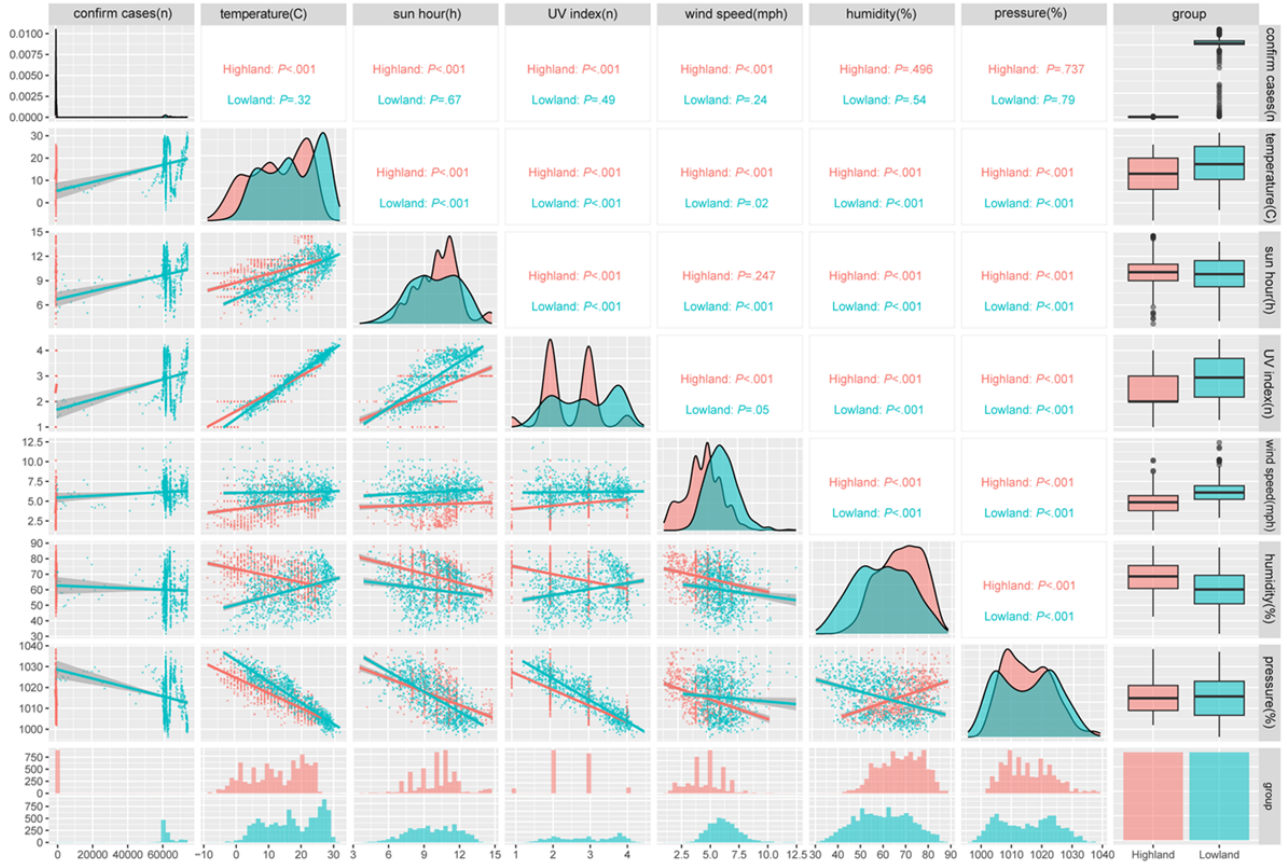
**Figure 1.** Pattern of confirmed COVID-19 cases at high altitudes within the study period. (A) Geographic patterns of confirmed COVID-19 cases in China as of July 7, 2022. (B) Comparison of confirmed COVID-19 cases reported in different cities at different altitudes. (C) Linear regression analysis 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 (ρ=0.388, $P$<.001) was positively correlated with confirmed cases.

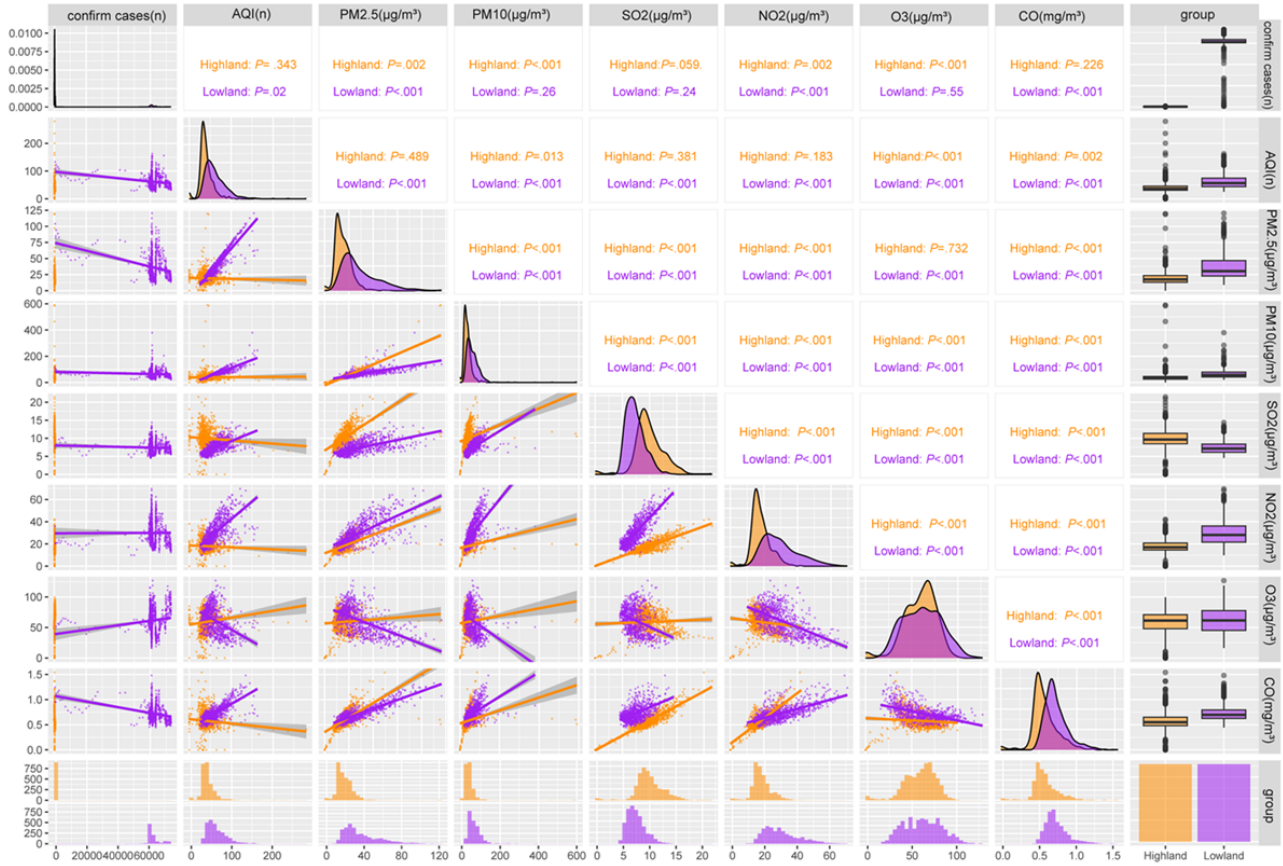
**Figure 2.** Spearman correlation analysis of confirmed COVID-19 cases and meteorological factors in high-altitude and low-altitude regions in China. The color represents each region group. Area charts represent frequencies and the scatterplots with lines represent the correlation. *P* values are based 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.

**Figure 3.** Spearman correlation analysis of confirmed COVID-19 cases and air quality factors in high- and low-altitude regions in China. The color represents each region group. Area charts represent frequencies and the scatterplot with lines represent correlations. AQI: air quality index; CO: carbon monoxide; NO2: nitrogen dioxide; O3: ozone; PM2.5: fine particulate matter; PM10: coarse particulate matter; SO2: sulfur dioxide. *P* values are based on 2-tailed tests; also see 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.

| Fixed effects | Model 1[a] | | Model 2[b] | | Model 3[c] | | Model 4[d] | |
|---|---|---|---|---|---|---|---|---|
| | Estimate (SE) | *P* value | Estimate (SE) | *P* value | Estimate (SE) | *P* value | Estimate (SE) | *P* value |
| Average temperature (°C) | –15.41 (2.55) | .001 | –21.44 (7.47) | .004 | —[e] | — | –27.90 (7.72) | .001 |
| Sunlight hours | –45.31 (9.50) | .001 | –12.55 (15.04) | .40 | — | — | –8.86 (15.31) | .56 |
| UV index | –201.24 (27.53) | .001 | –257.41 (71.06) | .001 | — | — | 249.03 (71.84) | .05 |
| Wind speed | 7.934 (10.81) | .46 | 16.72 (11.27) | .13 | — | — | 16.71 (12.00) | .02 |
| Humidity | –0.071 (1.42) | .96 | –1.99 (1.81) | .27 | — | — | 1.9 8 (1.94) | .31 |
| Pressure | 17.64 (2.51) | .001 | 20.35 (5.51) | .001 | — | — | 24.32 (5.58) | .05 |
| AQI[f] | –1.45 (0.67) | .03 | — | — | –0.05 (0.88) | .94 | 0.11 (0.89) | .90 |
| PM2.5[g] | 4.71 (0.99) | .001 | — | — | 4.28 (1.35) | .15 | 2.96 (1.36) | .05 |
| PM10[h] | 0.36 (0.47) | .43 | — | — | 0.49 (0.58) | .39 | 0.52 (0.58) | .36 |
| SO2[i] | 6.93 (6.01) | .24 | — | — | 18.41 (6.75) | .64 | 22.06 (6.94) | .14 |
| NO2[j] | 1.88 (1.612) | .24 | — | — | 8.69 (1.97) | .14 | 10.21 (2.07) | .07 |
| O3[k] | 2.29 (0.77) | .002 | — | — | 2.09 (0.8) | .008 | 0.56 (0.86) | .51 |
| CO[l] | 375.4 (75.5) | .001 | — | — | 471.52 (101.55) | .001 | 564.31 (107.99) | .001 |

[a]Model 1 includes each of the fixed effects run in 12 separate models.

[b]Model 2 includes all meteorological covariates considered simultaneously.

[c]Model 3 includes all air quality covariates considered simultaneously.

[d]Model 4 includes all 12 environmental indicators considered simultaneously.

[e]Excluded from model.

[f]AQI: air quality index.

[g]PM2.5: fine particulate matter.

[h]PM10: coarse particulate matter.

[i]SO2: sulfur dioxide.

[j]NO2: nitrogen dioxide.

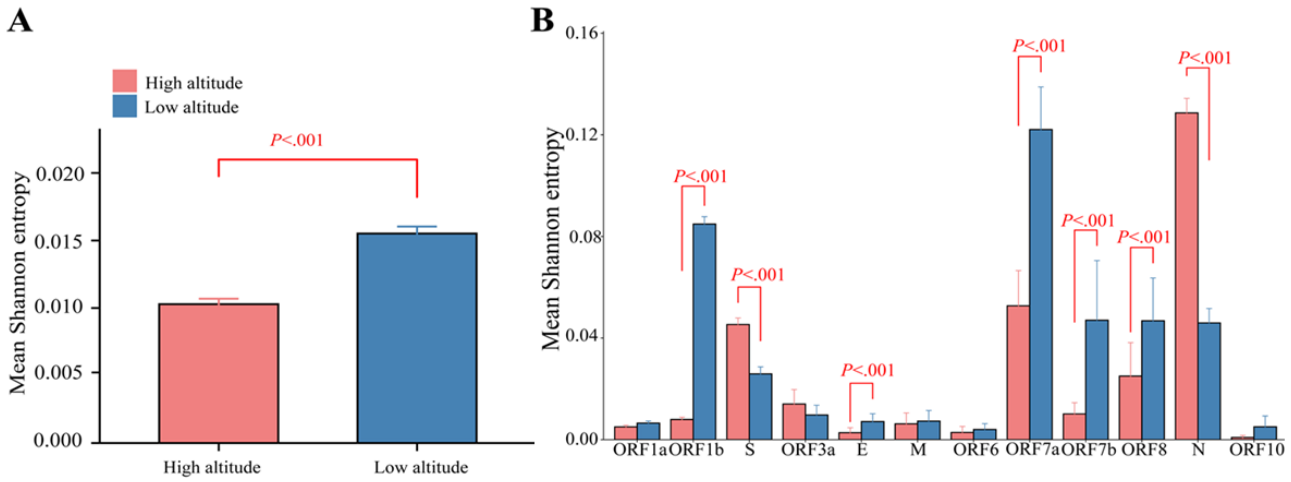
[k]O3: ozone.

[l]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.

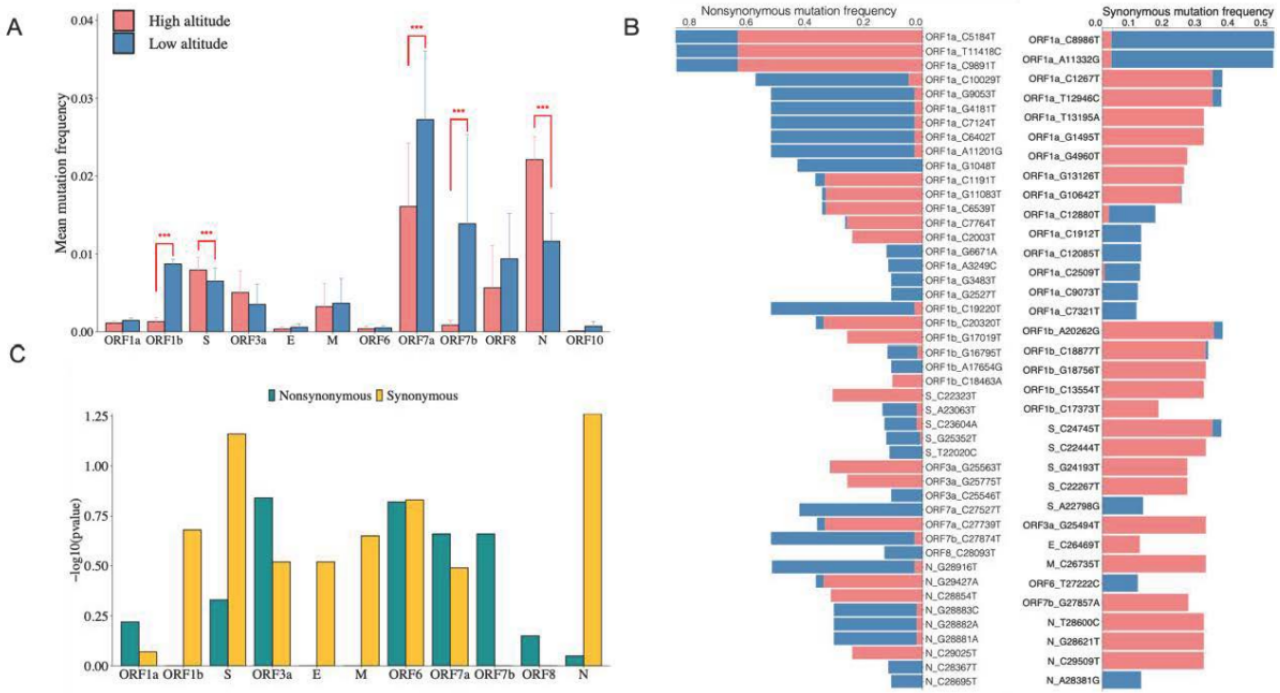
**Figure 4.** Mean sequence diversity across high-altitude (red) and low-altitude (blue) groups identified by Shannon entropy in nucleotides and different amino acids of the SARS-CoV-2 genome. (A) Mean sequence diversity identified by Shannon entropy in nucleotides of the whole SARS-CoV-2 genome. (B) 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.

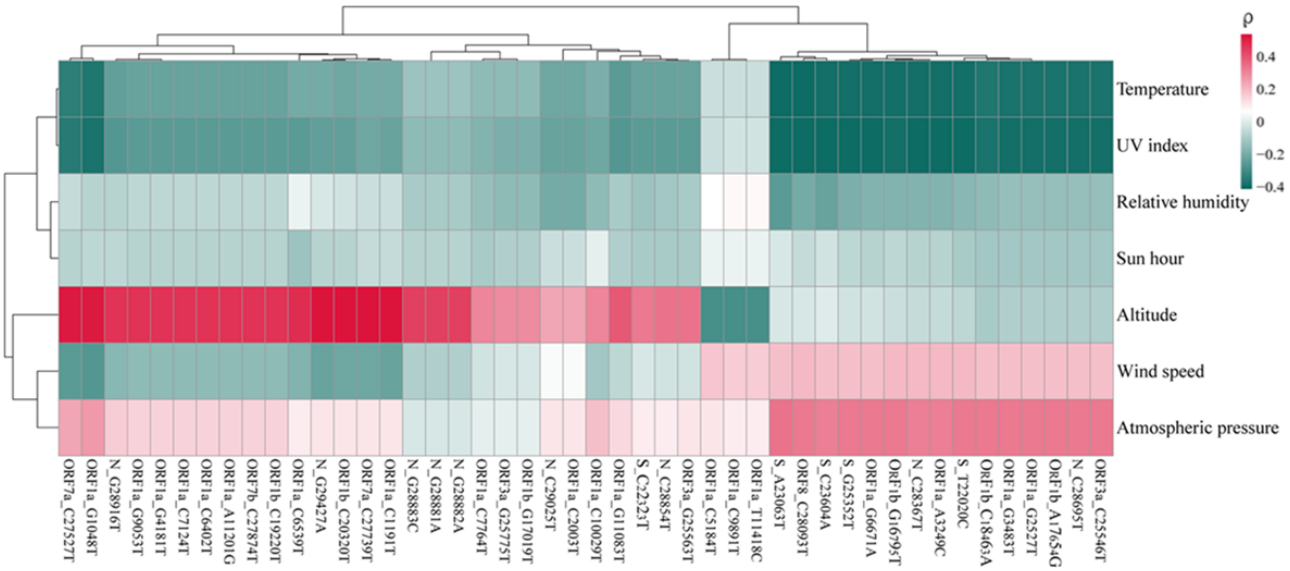
**Figure 5.** Distribution of mutations of SARS-CoV-2 in the high- and low-altitude groups. (A) Mean mutation frequency of open reading frames (ORFs) in the high- and low-altitude groups. (B) Comparison of frequencies of nonsynonymous and synonymous mutations between the high- and low-altitude groups. (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 (ρ=0.53, *P*<.001); followed by UV index with ORF8_C28093T and S_A23063T (ρ=–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 (ρ=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 (ρ=0.34 and ρ=0.31, respectively; both *P*<.001) and a significant negative correlation with temperature and UV index (ρ=–0.42 and ρ=–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 (ρ=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 S255F (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 ]

Multimedia Appendix 8
Spearman correlation analysis of environmental factors and mutation frequency in the high- and low-altitude groups.
[XLSX File (Microsoft Excel File), 17 KB - ijmr_v13i1e43585_app8.xlsx ]

Multimedia Appendix 9
Frequencies of nonsynonymous and synonymous mutations of the main lineages of SARS-CoV-2 in the high- and low-altitude groups.
[XLSX File (Microsoft Excel File), 10 KB - ijmr_v13i1e43585_app9.xlsx ]

Multimedia Appendix 10
Number of samples of different SARS-CoV-2 lineages in the high-altitude (red) and low-altitude (blue) regions.
[DOCX File , 64 KB - ijmr_v13i1e43585_app10.docx ]

## References


1. Zhang Z, Fang M, Wu R, Zong H, Huang H, Tong Y, et al. Large-scale biomedical relation extraction across diverse relation types: model development and usability study on COVID-19. J Med Internet Res 2023 Sep 20;25:e48115 [FREE Full text] [doi: 10.2196/48115] [Medline: 37632414]
2. Bashir MF, Ma B, Bilal, Komal B, Bashir MA, Tan D, et al. Correlation between climate indicators and COVID-19 pandemic in New York, USA. Sci Total Environ 2020 Aug 01;728:138835 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138835] [Medline: 32334162]
3. Yao Y, Pan J, Liu Z, Meng X, Wang W, Kan H, et al. No association of COVID-19 transmission with temperature or UV radiation in Chinese cities. Eur Respir J 2020 May;55(5):55 [FREE Full text] [doi: 10.1183/13993003.00517-2020] [Medline: 32269084]
4. Cano-Pérez E, Torres-Pacheco J, Fragozo-Ramos MC, García-Díaz G, Montalvo-Varela E, Pozo-Palacios JC. Negative correlation between altitude and COVID-19 pandemic in Colombia: a preliminary report. Am J Trop Med Hyg 2020 Dec;103(6):2347-2349 [FREE Full text] [doi: 10.4269/ajtmh.20-1027] [Medline: 33124543]
5. Chen H, Qin L, Wu S, Xu W, Gao R, Zhang X. Clinical characteristics and laboratory features of COVID-19 in high altitude areas: a retrospective cohort study. PLoS One 2021;16(5):e0249964 [FREE Full text] [doi: 10.1371/journal.pone.0249964] [Medline: 34003821]
6. Hakami AR, Dobie G. Studying the effect of particulate matter as SARS-CoV-2 transmitters. J Public Health Res 2021 Oct 21;11(1):2521 [FREE Full text] [doi: 10.4081/jphr.2021.2521] [Medline: 34674518]
7. Segovia-Juarez J, Castagnetto JM, Gonzales GF. High altitude reduces infection rate of COVID-19 but not case-fatality rate. Respir Physiol Neurobiol 2020 Oct;281:103494 [FREE Full text] [doi: 10.1016/j.resp.2020.103494] [Medline: 32679369]
8. Arias-Reyes C, Zubieta-DeUrioste N, Poma-Machicao L, Aliaga-Raduan F, Carvajal-Rodriguez F, Dutschmann M, et al. Does the pathogenesis of SARS-CoV-2 virus decrease at high-altitude? Respir Physiol Neurobiol 2020 Jun;277:103443 [FREE Full text] [doi: 10.1016/j.resp.2020.103443] [Medline: 32333993]
9. Ortiz-Prado E, Simbaña-Rivera K, Fernandez-Naranjo R, Vásconez JE, Henriquez-Trujillo AR, Vallejo-Janeta AP, UDLA-COVID-19 Team. SARS-CoV-2 viral load analysis at low and high altitude: a case study from Ecuador. Int J Environ Res Public Health 2022 Jun 28;19(13):7945 [FREE Full text] [doi: 10.3390/ijerph19137945] [Medline: 35805606]
10. Bhadra A, Mukherjee A, Sarkar K. Impact of population density on Covid-19 infected and mortality rate in India. Model Earth Syst Environ 2021;7(1):623-629 [FREE Full text] [doi: 10.1007/s40808-020-00984-7] [Medline: 33072850]
11. Ganasegeran K, Jamil MFA, Ch'ng ASH, Looi I, Peariasamy KM. Influence of population density for COVID-19 spread in Malaysia: an ecological study. Int J Environ Res Public Health 2021 Sep 18;18(18):9866 [FREE Full text] [doi: 10.3390/ijerph18189866] [Medline: 34574790]
12. Sun Z, Zhang H, Yang Y, Wan H, Wang Y. Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of China. Sci Total Environ 2020 Dec 01;746:141347 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.141347] [Medline: 32755746]
13. Sy KTL, White LF, Nichols BE. Population density and basic reproductive number of COVID-19 across United States counties. PLoS One 2021 Apr 21;16(4):e0249271 [FREE Full text] [doi: 10.1371/journal.pone.0249271] [Medline: 33882054]
14. Mengist HM, Kombe Kombe AJ, Mekonnen D, Abebaw A, Getachew M, Jin T. Mutations of SARS-CoV-2 spike protein: implications on immune evasion and vaccine-induced immunity. Semin Immunol 2021 Jun;55:101533 [FREE Full text] [doi: 10.1016/j.smim.2021.101533] [Medline: 34836774]
15. Ramesh S, Govindarajulu M, Parise RS, Neel L, Shankar T, Patel S, et al. Emerging SARS-CoV-2 variants: a review of its mutations, its implications and vaccine efficacy. Vaccines 2021 Oct 18;9(10):1195 [FREE Full text] [doi: 10.3390/vaccines9101195] [Medline: 34696303]

16. Tong Y, Tan F, Huang H, Zhang Z, Zong H, Xie Y, et al. ViMRT: a text-mining tool and search engine for automated virus mutation recognition. Bioinformatics 2023 Jan 01;39(1):btac721 [FREE Full text] [doi: 10.1093/bioinformatics/btac721] [Medline: 36342236]
17. Hao Z, Li R, Hao C, Zhao H, Wan X, Guo D. Global evidence of temperature acclimation of COVID-19 D614G linage. Glob Chall 2021 Jun;5(6):2000132 [FREE Full text] [doi: 10.1002/gch2.202000132] [Medline: 33786200]
18. Sharif N, Dey SK. Impact of population density and weather on COVID-19 pandemic and SARS-CoV-2 mutation frequency in Bangladesh. Epidemiol Infect 2021 Jan 07;149:e16 [FREE Full text] [doi: 10.1017/S0950268821000029] [Medline: 33407987]
19. Sharif N, Sarkar MK, Ahmed SN, Ferdous RN, Nobel NU, Parvez AK, et al. Environmental correlation and epidemiologic analysis of COVID-19 pandemic in ten regions in five continents. Heliyon 2021 Mar;7(3):e06576 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06576] [Medline: 33817387]
20. Yang T, Yu P, Chang Y, Hsu SD. D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation. J Biol Chem 2021 Oct;297(4):101238 [FREE Full text] [doi: 10.1016/j.jbc.2021.101238] [Medline: 34563540]
21. COVID-19 Global Pandemic Real-time Report. DXY-DX Doctor. URL: https://ncov.dxy.cn/ncovh5/view/en_pneumonia?from=dxy&source=&link=&share= [accessed 2024-03-13]
22. Topographic map. URL: https://en-gb.topographic-map.com/map-d821h/China/ [accessed 2024-03-13]
23. World Weather API and Weather Forecast. World Weather Online. URL: https://www.worldweatheronline.com/ [accessed 2024-03-13]
24. Global Initiative on Sharing All Influenza Data (GISAID). URL: https://www.gisaid.org/ [accessed 2024-03-13]
25. Zagordi O, Däumer M, Beisel C, Beerenwinkel N. Read length versus depth of coverage for viral quasispecies reconstruction. PLoS One 2012 Oct 3;7(10):e47046 [FREE Full text] [doi: 10.1371/journal.pone.0047046] [Medline: 23056573]
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. URL: https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm [accessed 2024-03-13]
28. Rocklöv J, Sjödin H. High population densities catalyse the spread of COVID-19. J Travel Med 2020 May 18;27(3):taaa038 [FREE Full text] [doi: 10.1093/jtm/taaa038] [Medline: 32227186]
29. Li H, Xu X, Dai D, Huang Z, Ma Z, Guan Y. Air pollution and temperature are associated with increased COVID-19 incidence: a time series study. Int J Infect Dis 2020 Aug;97:278-282 [FREE Full text] [doi: 10.1016/j.ijid.2020.05.076] [Medline: 32502664]
30. Pirouz B, Shaffiee Haghshenas S, Pirouz B, Shaffiee Haghshenas S, Piro P. Development of an assessment method for investigating the impact of climate and urban parameters in confirmed cases of COVID-19: a new challenge in sustainable development. Int J Environ Res Public Health 2020 Apr 18;17(8):2801 [FREE Full text] [doi: 10.3390/ijerph17082801] [Medline: 32325763]
31. Ujiie M, Tsuzuki S, Ohmagari N. Effect of temperature on the infectivity of COVID-19. Int J Infect Dis 2020 Jun;95:301-303 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.068] [Medline: 32360939]
32. Song P, Han H, Feng H, Hui Y, Zhou T, Meng W, et al. High altitude relieves transmission risks of COVID-19 through meteorological and environmental factors: evidence from China. Environ Res 2022 Sep;212(Pt B):113214 [FREE Full text] [doi: 10.1016/j.envres.2022.113214] [Medline: 35405128]
33. Carleton T, Cornetet J, Huybers P, Meng KC, Proctor J. Global evidence for ultraviolet radiation decreasing COVID-19 growth rates. Proc Natl Acad Sci U S A 2021 Jan 07;118(1):e2012370118 [FREE Full text] [doi: 10.1073/pnas.2012370118] [Medline: 33323525]
34. Ratnesar-Shumate S, Williams G, Green B, Krause M, Holland B, Wood S, et al. Simulated sunlight rapidly inactivates SARS-CoV-2 on surfaces. J Infect Dis 2020 Jun 29;222(2):214-222 [FREE Full text] [doi: 10.1093/infdis/jiaa274] [Medline: 32432672]
35. Pani SK, Lin N, RavindraBabu S. Association of COVID-19 pandemic with meteorological parameters over Singapore. Sci Total Environ 2020 Oct 20;740:140112 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.140112] [Medline: 32544735]
36. Fiasca F, Minelli M, Maio D, Minelli M, Vergallo I, Necozione S, et al. Associations between COVID-19 incidence rates and the exposure to PM2.5 and NO: a nationwide observational study in Italy. Int J Environ Res Public Health 2020 Dec 13;17(24):9318 [FREE Full text] [doi: 10.3390/ijerph17249318] [Medline: 33322089]
37. Isaia G, Diémoz H, Maluta F, Fountoulakis I, Ceccon D, di Sarra A, et al. Does solar ultraviolet radiation play a role in COVID-19 infection and deaths? An environmental ecological study in Italy. Sci Total Environ 2021 Feb 25;757:143757 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.143757] [Medline: 33272604]
38. Pansini R, Fornacca D. COVID-19 higher mortality in Chinese regions with chronic exposure to lower air quality. Front Public Health 2020;8:597753 [FREE Full text] [doi: 10.3389/fpubh.2020.597753] [Medline: 33585383]
39. Ismail IN, Jalaludin J, Bakar SA, Hisamuddin NH, Suhaimi NF. Association of traffic-related air pollution (TRAP) with DNA damage and respiratory health symptoms among primary school children in Selangor. Asian J Atmosph Sci 2019 Jun 30;13(2):106-116. [doi: 10.5572/ajae.2019.13.2.106]

40. Pansini R, Fornacca D. Early spread of COVID-19 in the air-polluted regions of eight severely affected countries. Atmosphere 2021 Jun 21;12(6):795. [doi: 10.3390/atmos12060795]
41. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
42. Xie J, Zhu Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci Total Environ 2020 Jul 01;724:138201 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138201] [Medline: 32408450]
43. Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and COVID-19 infection: evidence from China. Sci Total Environ 2020 Jul 20;727:138704 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138704] [Medline: 32315904]
44. Srivastava S, Garg I, Bansal A, Kumar B. SARS-CoV-2 infection: physiological and environmental gift factors at high altitude. Virusdisease 2020 Dec;31(4):450-452 [FREE Full text] [doi: 10.1007/s13337-020-00626-7] [Medline: 32953947]
45. Stephens KE, Chernyavskiy P, Bruns DR. Impact of altitude on COVID-19 infection and death in the United States: a modeling and observational study. PLoS One 2021;16(1):e0245055 [FREE Full text] [doi: 10.1371/journal.pone.0245055] [Medline: 33444357]
46. Sivasubramanian S, Gopalan V, Ramesh K, Padmanabhan P, Mone K, Govindan K, et al. Phylodynamic pattern of genetic clusters, paradigm shift on spatio-temporal distribution of clades, and impact of spike glycoprotein mutations of SARS-CoV-2 isolates from India. J Glob Infect Dis 2021;13(4):164-171 [FREE Full text] [doi: 10.4103/jgid.jgid_97_21] [Medline: 35017872]
47. Sun S, Gu H, Cao L, Chen Q, Ye Q, Yang G, et al. Characterization and structural basis of a lethal mouse-adapted SARS-CoV-2. Nat Commun 2021 Sep 27;12(1):5654. [doi: 10.1038/s41467-021-25903-x] [Medline: 34580297]
48. Mustafa Z, Kalbacher H, Burster T. Occurrence of a novel cleavage site for cathepsin G adjacent to the polybasic sequence within the proteolytically sensitive activation loop of the SARS-CoV-2 Omicron variant: The amino acid substitution N679K and P681H of the spike protein. PLoS One 2022;17(4):e0264723 [FREE Full text] [doi: 10.1371/journal.pone.0264723] [Medline: 35436320]
49. Dyer O. Covid-19: China stops counting cases as models predict a million or more deaths. BMJ 2023 Jan 03;380:2. [doi: 10.1136/bmj.p2] [Medline: 36596574]


## Abbreviations

**AQI:** air quality index
**CO:** carbon monoxide
**FP:** fusion peptide
**GISAID:** Global Initiative on Sharing All Influenza Data
**NO2:** nitrogen dioxide
**O3:** ozone
**ORF:** open reading frame
**PM2.5:** fine particulate matter
**PM10:** coarse particulate matter
**RBD:** receptor-binding domain
**Sn:** Shannon entropy


*Edited by T de Azevedo Cardoso; submitted 18.10.22; peer-reviewed by M Kapsetaki, R Menhas, Y Zhu, D Pala; comments to author 17.05.23; revised version received 20.07.23; accepted 06.03.24; published 25.03.24.*

*Please cite as:*
*Deji Z, Tong Y, Huang H, Zhang Z, Fang M, Crabbe MJC, Zhang X, Wang Y*
*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*
*URL: https://www.i-jmr.org/2024/1/e43585*
*doi:10.2196/43585*
*PMID:38526532*

Original Paper

# 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 Boparai[1], BSc; Sarah Dunnett[1], BA; Michelle Wu[1], BSc, MSc; Vanessa K Tassone[1], BA; Sophie F Duffy[1], BSc, MSc; Valentina Zuluaga Cuartas[1], BSc; Ziming Chen[1], BSc, MSc; Hyejung Jung[2], BSc, MSc; Catherine M Sabiston[3], BSc, MHK, PhD; Wendy Lou[2], PhD; Venkat Bhat[1,4,5,6], MSc, MD

[1]Interventional Psychiatry Program, St. Michael's Hospital, Toronto, ON, Canada

[2]Department of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Department of Kinesiology, Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, ON, Canada

[4]Mental Health and Addictions Services, St. Michael's Hospital, Toronto, ON, Canada

[5]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[6]Institute 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 (adjusted coefficient [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.

***(Interact J Med Res 2024;13:e48396)*** doi:10.2196/48396

**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/m$^2$ [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 Centre 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.

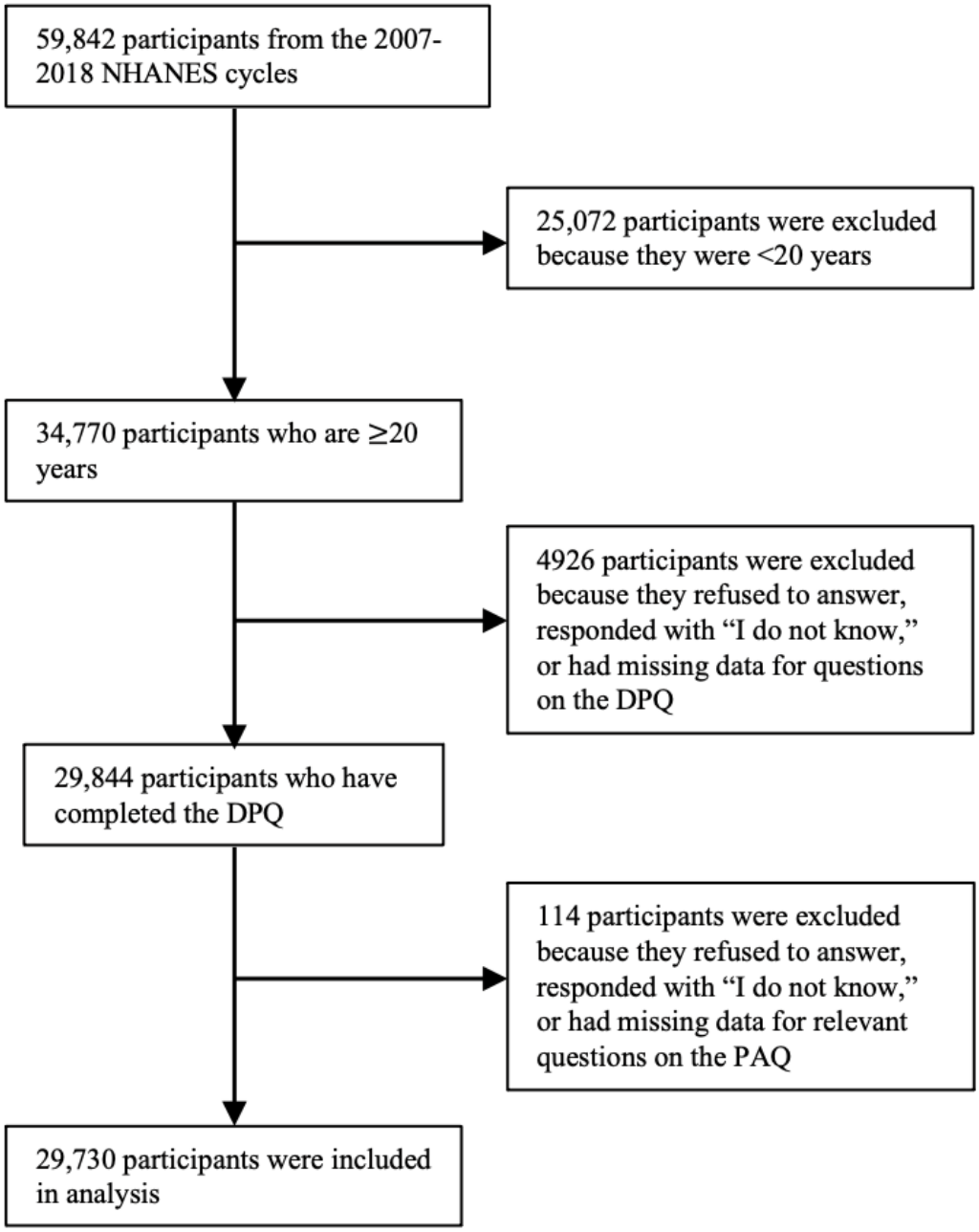


**Table 1.** Characteristics of included participants in the cross-sectional analyses from a nationally representative sample of the United States obtained through the National Health and Nutrition Examination Survey, 2007-2018[a].

| Characteristics | No depressive symptoms (<10; n=26,991) | Depressive symptoms (≥10; n=2739) | *P* value |
|---|---|---|---|
| Age (years), mean (SD) | 47.68 (17.08) | 46.99 (15.94) | .15 |
| **Sex, n (%)** | | | <.001 |
| Female | 13,385 (50.18) | 1748 (64.54) | |
| Male | 13,606 (49.82) | 991 (35.46) | |
| **Race, n (%)** | | | <.001 |
| Mexican American | 4059 (8.52) | 405 (8.03) | |
| Non-Hispanic Black | 5775 (10.85) | 588 (13.27) | |
| Non-Hispanic White | 11,210 (67.39) | 1157 (63.31) | |
| Other Hispanic | 2724 (5.59) | 377 (8.01) | |
| Other race or multiracial | 3223 (7.66) | 212 (7.38) | |
| **Education, n (%)** | | | <.001 |
| Less than high school | 6107 (14.41) | 979 (25.51) | |
| High school or GED[b] | 6173 (22.82) | 653 (26.54) | |
| Some college | 8047 (31.4) | 819 (33.78) | |
| College and above | 6643 (31.37) | 285 (14.17) | |
| **Marital status, n (%)** | | | <.001 |
| Married or living with other | 16,366 (64.59) | 1208 (46.9) | |
| Divorced or separated or widowed | 5718 (17.34) | 948 (31.22) | |
| Never married | 4892 (18.07) | 580 (21.88) | |
| **PIR[c], n (%)** | | | <.001 |
| ≤1.3 (low income) | 7335 (19.84) | 1319 (41.89) | |
| >1.3 (mid-to-high income) | 17,248 (80.16) | 1153 (58.11) | |
| **BMI (kg/m$^2$), n (%)** | | | <.001 |
| <18 | 392 (1.42) | 50 (2.12) | |
| 18 to <25 | 7259 (27.93) | 590 (23.9) | |
| 25 to <30 | 8975 (33.49) | 708 (26.09) | |
| ≥30 | 10,112 (37.16) | 1350 (47.89) | |
| **Cigarette use, n (%)** | | | <.001 |
| Yes | 5054 (17.85) | 1014 (39.37) | |
| No | 21,920 (82.15) | 1725 (60.63) | |
| **Health condition, n (%)** | | | <.001 |
| Good or above | 21,403 (85.17) | 1111 (49.08) | |
| Fair | 4989 (13.33) | 1111 (36.19) | |
| Poor | 597 (1.50) | 517 (14.73) | |
| **PA[d] (hours per week), mean (SD)** | | | |
| Total PA | 10.87 (16.85) | 8.82 (16.27) | <.001 |
| Moderate-intensity PA | 6.81 (10.98) | 5.71 (10.83) | .001 |
| Vigorous-intensity PA | 4.06 (9.27) | 3.10 (8.50) | <.001 |
| Recreational PA | 2.72 (4.46) | 1.52 (4.02) | <.001 |
| Work-related PA | 8.15 (15.83) | 7.30 (15.19) | .06 |

| Characteristics | No depressive symptoms (<10; n=26,991) | Depressive symptoms (≥10; n=2739) | *P* value |
|---|---|---|---|
| Moderate-intensity work-related PA | 5.17 (10.27) | 4.64 (9.98) | .07 |
| Vigorous-intensity work-related PA | 2.98 (8.73) | 2.66 (8.09) | .17 |
| Moderate-intensity recreational PA | 1.63 (3.15) | 1.08 (3.09) | <.001 |
| Vigorous-intensity recreational PA | 1.08 (2.63) | 0.44 (1.92) | <.001 |
| PHQ-9[e], mean (SD) | 2.13 (2.39) | 14.06 (3.81) | <.001 |
| Cognitive-affective, mean (SD) | 0.78 (1.34) | 7.85 (3.27) | <.001 |
| Somatic, mean (SD) | 1.35 (1.55) | 6.21 (1.94) | <.001 |

[a]Unweighted frequencies are paired with weighted percentages. Mean and SDs are weighted. *P* values are based on weighted values.

[b]GED: General Educational Development.

[c]PIR: poverty-income ratio.

[d]PA: physical activity.

[e]PHQ-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[a] and aORs[b] for associations between PA[c] 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[d].

| PA | Depressive symptoms (PHQ-9[e]≥10) | | | |
|---|---|---|---|---|
| | OR (95% CI) | *P* value | aOR (95% CI) | *P* value |
| Total PA | 0.992 (0.987-0.996) | .002 | 0.996 (0.991-1.000) | .20 |
| Moderate-intensity PA | 0.990 (0.983-0.997) | .007 | 0.996 (0.989-1.003) | .54 |
| Vigorous-intensity PA | 0.986 (0.978-0.995) | .005 | 0.991 (0.983-0.999) | .14 |
| Recreational PA | 0.902 (0.877-0.928) | <.001 | 0.965 (0.944-0.986) | .007 |
| Work-related PA | 0.996 (0.992-1.000) | .08 | 0.997 (0.993-1.002) | .54 |

[a]OR: 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 3.** Associations between PA[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[b].

| PA | Somatic symptoms | | Cognitive-affective symptoms | |
|---|---|---|---|---|
| | aβ[c] (95% CI) | *P* value | aβ (95% CI) | *P* value |
| Total PA | –0.001 (–0.003 to 0.001) | ≥.99 | –0.002 (–0.005 to 0.000) | .12 |
| Moderate-intensity PA | 0.000 (–0.004 to 0.004) | ≥.99 | –0.002 (–0.005 to 0.001) | .36 |
| Vigorous-intensity PA | –0.003 (–0.006 to 0.001) | .41 | –0.004 (–0.008 to 0.000) | .12 |
| Recreational PA | –0.016 (–0.022 to –0.009) | <.001 | –0.015 (–0.023 to –0.007) | .003 |
| Work-related PA | 0.000 (–0.002 to 0.003) | ≥.99 | –0.001 (–0.004 to 0.001) | .36 |

[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.

### 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 [aβ]=–0.021, 95% CI –0.032 to –0.010; *P*=.001) and cognitive-affective (aβ=–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 (aβ=–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).

**Table 4.** Weighted ORs[a] and aORs[b] for associations between PA[c] 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[d].

| PA | Depressive symptoms (PHQ-9[e]≥10) | | | |
|---|---|---|---|---|
| | OR (95% CI) | *P* value | aOR (95% CI) | *P* value |
| Moderate-intensity work-related PA | 0.995 (0.988-1.001) | .19 | 0.998 (0.990-1.005) | .53 |
| Vigorous-intensity work-related PA | 0.995 (0.989-1.002) | .19 | 0.995 (0.987-1.002) | .35 |
| Moderate-intensity recreational PA | 0.916 (0.882-0.952) | <.001 | 0.975 (0.949-1.001) | .19 |
| Vigorous-intensity recreational PA | 0.829 (0.783-0.879) | <.001 | 0.926 (0.883-0.972) | .009 |

[a]OR: 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[a] 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[b].

| PA | Somatic symptoms | | Cognitive-affective symptoms | |
|---|---|---|---|---|
| | aβ[c] (95% CI) | *P* value | aβ (95% CI) | *P* value |
| Moderate-intensity work-related PA | 0.002 (–0.002 to 0.006) | .79 | –0.001 (–0.004 to 0.002) | .55 |
| Vigorous-intensity work-related PA | –0.001 (–0.005 to 0.003) | .79 | –0.003 (–0.007 to 0.001) | .34 |
| Moderate-intensity recreational PA | –0.016 (–0.025 to –0.007) | .002 | –0.014 (–0.026 to –0.003) | .05 |
| Vigorous-intensity recreational PA | –0.021 (–0.032 to –0.010) | .001 | –0.022 (–0.035 to –0.009) | .005 |

[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.

# Discussion

## Principal Findings

This study investigated the association of the weekly duration of PA (including subsets by intensity, domain, or both) with the overall presence of depressive symptoms and somatic and cognitive-affective depression symptomatology. The results of this study indicate that it is the PA domain, specifically recreational PA, that may be crucial to the relationship between 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

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.
[DOCX File , 25 KB - ijmr_v13i1e48396_app1.docx ]

## References

1. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2024-08-04]
2. Fox KR. The influence of physical activity on mental well-being. Public Health Nutr 1999;2:411-418. [doi: 10.1017/s1368980099000567] [Medline: 10610081]
3. Craft LL, Landers DM. The effect of exercise on clinical depression and depression resulting from mental illness: a meta-analysis. J Sport Exerc Psychol 1998;20(4):339-357. [doi: 10.1123/jsep.20.4.339]
4. Conn VS. Depressive symptom outcomes of physical activity interventions: meta-analysis findings. Ann Behav Med 2010;39(2):128-138 [FREE Full text] [doi: 10.1007/s12160-010-9172-x] [Medline: 20422333]
5. Dimeo F, Bauer M, Varahram I, Proest G, Halter U. Benefits from aerobic exercise in patients with major depression: a pilot study. Br J Sports Med 2001;35(2):114-117 [FREE Full text] [doi: 10.1136/bjsm.35.2.114] [Medline: 11273973]
6. Camacho TC, Roberts RE, Lazarus NB, Kaplan GA, Cohen RD. Physical activity and depression: evidence from the Alameda County Study. Am J Epidemiol 1991;134(2):220-231. [doi: 10.1093/oxfordjournals.aje.a116074] [Medline: 1862805]
7. Physical activity guidelines for Americans. 2nd Edition. US Department of Health and Human Services. 2018. URL: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2024-06-08]
8. Motl RW, Konopack JF, McAuley E, Elavsky S, Jerome GJ, Marquez DX. Depressive symptoms among older adults: long-term reduction after a physical activity intervention. J Behav Med 2005;28(4):385-394. [doi: 10.1007/s10865-005-9005-5] [Medline: 16049630]
9. Guo F, Tian Y, Zhong F, Wu C, Cui Y, Huang C. Intensity of physical activity and depressive symptoms in college students: Fitness Improvement Tactics in Youth (FITYou) project. Psychol Res Behav Manag 2020;13:787-796 [FREE Full text] [doi: 10.2147/PRBM.S267066] [Medline: 33061694]

10. Wang X, Cai ZD, Jiang WT, Fang YY, Sun WX, Wang X. Systematic review and meta-analysis of the effects of exercise on depression in adolescents. Child Adolesc Psychiatry Ment Health 2022;16(1):16 [FREE Full text] [doi: 10.1186/s13034-022-00453-2] [Medline: 35227300]
11. Pearce M, Garcia L, Abbas A, Strain T, Schuch FB, Golubic R, et al. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry 2022;79(6):550-559 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.0609] [Medline: 35416941]
12. Hamer M, Stamatakis E, Steptoe A. Dose-response relationship between physical activity and mental health: the Scottish Health Survey. Br J Sports Med 2009;43(14):1111-1114. [doi: 10.1136/bjsm.2008.046243] [Medline: 18403415]
13. Hallgren M, Nguyen TTD, Owen N, Stubbs B, Vancampfort D, Lundin A, et al. Cross-sectional and prospective relationships of passive and mentally active sedentary behaviours and physical activity with depression. Br J Psychiatry 2020;217(2):413-419. [doi: 10.1192/bjp.2019.60] [Medline: 30895922]
14. Chen LJ, Stevinson C, Ku PW, Chang YK, Chu DC. Relationships of leisure-time and non-leisure-time physical activity with depressive symptoms: a population-based study of Taiwanese older adults. Int J Behav Nutr Phys Act 2012;9:28 [FREE Full text] [doi: 10.1186/1479-5868-9-28] [Medline: 22413813]
15. Pickett K, Yardley L, Kendrick T. Physical activity and depression: a multiple mediation analysis. Ment Health Phys Act 2012;5(2):125-134. [doi: 10.1016/j.mhpa.2012.10.001]
16. Teychenne M, Ball K, Salmon J. Physical activity and likelihood of depression in adults: a review. Prev Med 2008;46(5):397-411. [doi: 10.1016/j.ypmed.2008.01.009] [Medline: 18289655]
17. Mumba MN, Nacarrow AF, Cody S, Key BA, Wang H, Robb M, et al. Intensity and type of physical activity predicts depression in older adults. Aging Ment Health 2021;25(4):664-671. [doi: 10.1080/13607863.2020.1711861] [Medline: 31948269]
18. McKercher CM, Schmidt MD, Sanderson KA, Patton GC, Dwyer T, Venn AJ. Physical activity and depression in young adults. Am J Prev Med 2009;36(2):161-164. [doi: 10.1016/j.amepre.2008.09.036] [Medline: 19062235]
19. Kilpatrick M, Sanderson K, Blizzard L, Teale B, Venn A. Cross-sectional associations between sitting at work and psychological distress: reducing sitting time may benefit mental health. Ment Health Phys Act 2013;6(2):103-109. [doi: 10.1016/j.mhpa.2013.06.004]
20. Cerin E, Leslie E, Sugiyama T, Owen N. Associations of multiple physical activity domains with mental well-being. Ment Health Phys Act 2009;2(2):55-64. [doi: 10.1016/j.mhpa.2009.09.004]
21. Chilcot J, Rayner L, Lee W, Price A, Goodwin L, Monroe B, et al. The factor structure of the PHQ-9 in palliative care. J Psychosom Res 2013;75(1):60-64. [doi: 10.1016/j.jpsychores.2012.12.012] [Medline: 23751240]
22. Petersen JJ, Paulitsch MA, Hartig J, Mergenthal K, Gerlach FM, Gensichen J. Factor structure and measurement invariance of the Patient Health Questionnaire-9 for female and male primary care patients with major depression in Germany. J Affect Disord 2015;170:138-142. [doi: 10.1016/j.jad.2014.08.053] [Medline: 25240840]
23. Zipf G, Chiappa M, Porter KS, Ostchega Y, Lewis BG, Dostal J. National health and nutrition examination survey: plan and operations, 1999-2010. Vital Health Stat 1 2013(56):1-37 [FREE Full text] [Medline: 25078429]
24. Data User Agreement. Centers for Disease Control and Prevention. National Center for Health Statistics. URL: https://www.cdc.gov/nchs/data_access/restrictions.htm [accessed 2024-11-04]
25. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
26. Doi S, Ito M, Takebayashi Y, Muramatsu K, Horikoshi M. Factorial validity and invariance of the Patient Health Questionnaire (PHQ)-9 among clinical and non-clinical populations. PLoS One 2018;13(7):e0199235 [FREE Full text] [doi: 10.1371/journal.pone.0199235] [Medline: 30024876]
27. Zhang J, Yen ST. Physical activity, gender difference, and depressive symptoms. Health Serv Res 2015;50(5):1550-1573 [FREE Full text] [doi: 10.1111/1475-6773.12285] [Medline: 25630931]
28. Altemus M, Sarvaiya N, Neill Epperson C. Sex differences in anxiety and depression clinical perspectives. Front Neuroendocrinol 2014;35(3):320-330 [FREE Full text] [doi: 10.1016/j.yfrne.2014.05.004] [Medline: 24887405]
29. Bailey R, Mokonogho J, Kumar A. Racial and ethnic differences in depression: current perspectives. Neuropsychiatr Dis Treat 2019;15:603-609 [FREE Full text] [doi: 10.2147/NDT.S128584] [Medline: 30863081]
30. King AC, Blair SN, Bild DE, Dishman RK, Dubbert PM, Marcus BH, et al. Determinants of physical activity and interventions in adults. Med Sci Sports Exerc 1992;24(6 Suppl):S221-S236. [doi: 10.1249/00005768-199206001-00005] [Medline: 1625548]
31. Parks SE, Housemann RA, Brownson RC. Differential correlates of physical activity in urban and rural adults of various socioeconomic backgrounds in the United States. J Epidemiol Community Health 2003;57(1):29-35 [FREE Full text] [doi: 10.1136/jech.57.1.29] [Medline: 12490645]
32. Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BWJH, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry 2010;67(3):220-229. [doi: 10.1001/archgenpsychiatry.2010.2] [Medline: 20194822]
33. Tremblay MS, Ross R. How should we move for health? The case for the 24-hour movement paradigm. CMAJ 2020;192(49):E1728-E1729 [FREE Full text] [doi: 10.1503/cmaj.202345] [Medline: 33288508]

34. Fluharty M, Taylor AE, Grabski M, Munafò MR. The association of cigarette smoking with depression and anxiety: a systematic review. Nicotine Tob Res 2017;19(1):3-13 [FREE Full text] [doi: 10.1093/ntr/ntw140] [Medline: 27199385]
35. Policy basics: the Supplemental Nutrition Assistance Program (SNAP). Center on Budget and Policy Priorities. 2022. URL: https://www.cbpp.org/research/food-assistance/the-supplemental-nutrition-assistance-program-snap [accessed 2024-02-04]
36. About adult BMI. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html [accessed 2024-02-04]
37. Wade TJ, Cairney J, Pevalin DJ. Emergence of gender differences in depression during adolescence: national panel results from three countries. J Am Acad Child Adolesc Psychiatry 2002;41(2):190-198. [doi: 10.1097/00004583-200202000-00013] [Medline: 11837409]
38. Dijkstra-Kersten SMA, Biesheuvel-Leliefeld KEM, van der Wouden JC, Penninx BWJH, van Marwijk HWJ. Associations of financial strain and income with depressive and anxiety disorders. J Epidemiol Community Health 2015;69(7):660-665. [doi: 10.1136/jech-2014-205088] [Medline: 25636322]
39. Craft LL, Perna FM. The benefits of exercise for the clinically depressed. Prim Care Companion J Clin Psychiatry 2004;6(3):104-111 [FREE Full text] [doi: 10.4088/pcc.v06n0301] [Medline: 15361924]
40. Kandola A, Ashdown-Franks G, Hendrikse J, Sabiston CM, Stubbs B. Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity. Neurosci Biobehav Rev 2019;107:525-539. [doi: 10.1016/j.neubiorev.2019.09.040] [Medline: 31586447]
41. Lubans D, Richards J, Hillman C, Faulkner G, Beauchamp M, Nilsson M, et al. Physical activity for cognitive and mental health in youth: a systematic review of mechanisms. Pediatrics 2016;138(3):e20161642. [doi: 10.1542/peds.2016-1642] [Medline: 27542849]
42. Ekkekakis P, Hall EE, Petruzzello SJ. Variation and homogeneity in affective responses to physical activity of varying intensities: an alternative perspective on dose-response based on evolutionary considerations. J Sports Sci 2005;23(5):477-500. [doi: 10.1080/02640410400021492] [Medline: 16194996]
43. Goodman WK, Geiger AM, Wolf JM. Differential links between leisure activities and depressive symptoms in unemployed individuals. J Clin Psychol 2016;72(1):70-78 [FREE Full text] [doi: 10.1002/jclp.22231] [Medline: 26443982]
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]
45. Parnabas A, Mea KK, Parnabas J, Parnabas AM, Abdullah NM. Motives of outdoor recreational activities. : Springer; 2015 Presented at: Proceedings of the 2nd International Colloquium on Sports Science, Exercise, Engineering and Technology 2015; 2015; Singapore p. 215-223. [doi: 10.1007/978-981-287-691-1_23]
46. Byambasukh O, Snieder H, Corpeleijn E. Relation between leisure time, commuting, and occupational physical activity with blood pressure in 125 402 adults: the lifelines cohort. J Am Heart Assoc 2020;9(4):e014313 [FREE Full text] [doi: 10.1161/JAHA.119.014313] [Medline: 32067583]
47. Clays E, Hallman D, Oakman J, Holtermann A. Objectively measured occupational physical activity in blue-collar workers: what is the role of job type, gender and psychosocial resources? Appl Ergon 2020;82:102948. [doi: 10.1016/j.apergo.2019.102948] [Medline: 31493536]
48. White RL, Bennie J, Abbott G, Teychenne M. Work-related physical activity and psychological distress among women in different occupations: a cross-sectional study. BMC Public Health 2020;20(1):1007 [FREE Full text] [doi: 10.1186/s12889-020-09112-7] [Medline: 32586295]
49. Marttila J, Laitakari J, Nupponen R, Miilunpalo S, Paronen O. The versatile nature of physical activity—on the psychological, behavioural and contextual characteristics of health-related physical activity. Patient Educ Couns 1998;33(1 Suppl):S29-S38. [doi: 10.1016/s0738-3991(98)00007-x] [Medline: 10889744]
50. Wiles NJ, Haase AM, Lawlor DA, Ness A, Lewis G. Physical activity and depression in adolescents: cross-sectional findings from the ALSPAC cohort. Soc Psychiatry Psychiatr Epidemiol 2012;47(7):1023-1033 [FREE Full text] [doi: 10.1007/s00127-011-0422-4] [Medline: 21826444]
51. Goodwin RD. Association between physical activity and mental disorders among adults in the United States. Prev Med 2003;36(6):698-703. [doi: 10.1016/s0091-7435(03)00042-2] [Medline: 12744913]
52. West J, Otte C, Geher K, Johnson J, Mohr DC. Effects of Hatha yoga and African dance on perceived stress, affect, and salivary cortisol. Ann Behav Med 2004;28(2):114-118. [doi: 10.1207/s15324796abm2802_6] [Medline: 15454358]
53. Wise LA, Adams-Campbell LL, Palmer JR, Rosenberg L. Leisure time physical activity in relation to depressive symptoms in the Black women's health study. Ann Behav Med 2006;32(1):68-76 [FREE Full text] [doi: 10.1207/s15324796abm3201_8] [Medline: 16827631]
54. Lindwall M, Rennemark M, Halling A, Berglund J, Hassmén P. Depression and exercise in elderly men and women: findings from the Swedish national study on aging and care. J Aging Phys Act 2007;15(1):41-55. [doi: 10.1123/japa.15.1.41] [Medline: 17387228]
55. Oberste M, Medele M, Javelle F, Lioba Wunram H, Walter D, Bloch W, et al. Physical activity for the treatment of adolescent depression: a systematic review and meta-analysis. Front Physiol 2020;11:185 [FREE Full text] [doi: 10.3389/fphys.2020.00185] [Medline: 32265725]

56. Duclos M, Tabarin A. Exercise and the hypothalamo-pituitary-adrenal axis. Front Horm Res 2016;47:12-26. [doi: 10.1159/000445149] [Medline: 27348531]
57. Carek PJ, Laibstain SE, Carek SM. Exercise for the treatment of depression and anxiety. Int J Psychiatry Med 2011;41(1):15-28. [doi: 10.2190/PM.41.1.c] [Medline: 21495519]
58. Woo M, Kim S, Kim J, Petruzzello SJ, Hatfield BD. Examining the exercise-affect dose-response relationship: does duration influence frontal EEG asymmetry? Int J Psychophysiol 2009;72(2):166-172. [doi: 10.1016/j.ijpsycho.2008.12.003] [Medline: 19110012]
59. Stavrakakis N, de Jonge P, Ormel J, Oldehinkel AJ. Bidirectional prospective associations between physical activity and depressive symptoms. The TRAILS study. J Adolesc Health 2012;50(5):503-508. [doi: 10.1016/j.jadohealth.2011.09.004] [Medline: 22525115]
60. Azevedo Da Silva M, Singh-Manoux A, Brunner EJ, Kaffashian S, Shipley MJ, Kivimäki M, et al. Bidirectional association between physical activity and symptoms of anxiety and depression: the Whitehall II study. Eur J Epidemiol 2012;27(7):537-546 [FREE Full text] [doi: 10.1007/s10654-012-9692-8] [Medline: 22623145]
61. National Health and Nutrition Examination Survey. National Center for Health Statistics. 2024. URL: https://www.cdc.gov/nchs/nhanes/index.htm [accessed 2024-06-15]


## Abbreviations

**aβ:** 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 Azevedo 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:*
*Boparai JK, Dunnett S, Wu M, Tassone VK, Duffy SF, Zuluaga Cuartas V, Chen Z, Jung H, Sabiston CM, Lou W, Bhat V*
*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*
*Interact J Med Res 2024;13:e48396*
*URL: https://www.i-jmr.org/2024/1/e48396*
*doi:10.2196/48396*
*PMID:*

Original Paper

# Health and Well-Being Among College Students in the United States During the COVID-19 Pandemic: Daily Diary Study

Stephanie T Lanza[1], BA, BS, MS, MAS, PhD; Courtney Whetzel[2], BSc, MSc, PhD; Sandesh Bhandari[2], BS, MAS

[1]Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, United States

[2]The Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University, University Park, PA, United States

**Corresponding Author:**
Stephanie T Lanza, BA, BS, MS, MAS, PhD
Department of Biobehavioral Health
The Pennsylvania State University
302 Biobehavioral Health Building
University Park, PA, 16802
United States
Phone: 1 814 865 7095
Email: SLanza@psu.edu

## Abstract

**Background:** There is evidence that anxiety and stress increased among college students during the COVID-19 pandemic. However, less is known about daily experiences of affect, worry, substance use behaviors, experiences of pleasure, concern over food security, experiences of bias or discrimination, feelings of belongingness, and other indicators of well-being and how they vary across days in this population.

**Objective:** This study surveyed a wide range of indicators of health and well-being in daily life over 21 days with a sample of college students in a large university system in the United States during the pandemic. The overall variance in each daily measure was partitioned to reflect the proportion due to (1) between-person differences versus (2) within-person, day-to-day variability. This is important because measures that vary primarily due to between-person differences may be more amenable to interventions that target particular students, whereas measures that vary more due to day-to-day variability may be more amenable to interventions that target day-level contextual factors.

**Methods:** A sample of 2068 young adult college students (aged 18-24, mean 19.8, SD 1.3 years; 66.6% women) completed a baseline survey; 97.3% (n=2012) then completed up to 21 consecutive daily surveys that assessed a comprehensive set of daily markers of health, behavior, and well-being. The daily diary study produced a total of 33,722 person-days.

**Results:** Among all person-days, a minority were substance use days (eg, 14.5% of days involved alcohol use, 5.6% vaping, and 5.5% cannabis). Experiences of pleasure were reported on most (73.5%) days. Between-person differences explained more than 50% of the variance in numerous indicators of health and well-being, including daily vaping, cannabis use, other illicit substance use, experiences of bias or discrimination, positive affect, negative affect, worry, food insecurity, and feelings of belonging at the university. In contrast, within-person differences explained more than 50% of the variance in daily alcohol use, cigarette use, stress, experiences of pleasure, where the student slept last night, and physical activity.

**Conclusions:** College student health and well-being are multifaceted, with some aspects likely driven by person-level characteristics and experiences and other aspects by more dynamic, contextual risk factors that occur in daily life. These findings implicate services and interventions that should target individual students versus those that should target days on which students are at high risk for poor experiences or behaviors.

***(Interact J Med Res 2024;13:e45689)*** doi:10.2196/45689

**KEYWORDS**

daily diary; college student; young adult; mental health; substance use; stress; well-being; belonging

## Introduction

College students' well-being is multifaceted, including mental health, substance use behavior, stress, belonging, flourishing, and more. In terms of mental health, anxiety has been the most common problem reported by students [1]. Further, the prevalence of many substance use risk behaviors peaks during college ages, including 35% of young adults reporting past 30-day binge drinking (4+ and 5+ drinks in a row for females and males, respectively) [2]. This population also experiences high rates of acute consequences from alcohol use, including sexual assault and blacking out [3].

Much recent evidence shows that mental health problems among college students increased substantially during the COVID-19 pandemic [4,5]. However, while social isolation, anxiety, and depression increased during the pandemic, certain factors, such as a sense of belonging with their college, seem to be protective against the effects of the pandemic on key health outcomes, particularly among underrepresented racial or ethnic minority and first-generation students [6]. When considering college student health and well-being more broadly, a longitudinal study during the pandemic identified a large subgroup of students with either very low mental and behavioral health risk or elevated mental health symptoms but no other risks (eg, low rates of substance use, sexual activity, physical inactivity, and food insecurity) and a set of 4 smaller subgroups of students with particular configurations of problematic health and well-being, such as the intersection of elevated depression symptoms, recent cannabis use, and recent sexual activity [7].

A more intensive surveillance study of college students can shed important light on the dynamic nature of health and well-being by elucidating both individual characteristics and dynamic, contextual factors that may be linked to multiple facets of student physical and mental health and well-being. Toward this goal, the aim of the present study was to conduct a large daily diary study of college student experiences at the main campus of a large university in the Northeastern United States throughout the 2021 calendar year when the pandemic was still of great concern but students had fully returned to campus for primarily in-person learning. We sought to determine whether a broad range of daily indicators of health and well-being were attributable more to between-person differences or to day-to-day variability, which might indicate greater sensitivity to dynamic and contextualized experiences.

## Methods

### Recruitment and General Procedures

Data for this observational study were collected via web-based surveys from students at a large public university in spring 2021 and fall 2021 as part of the Student Engagement, Learning, and Flourishing study (ie, Project SELF). Over 35,000 undergraduate participants were invited to take part in the study. The inclusion criteria were that participants were full-time undergraduate students at this campus and aged 18-24 years.

Participants were sent the first recruitment email on a Tuesday, with a follow-up (if they had not yet completed the baseline survey) 1 week later. Undergraduate emails were separated into batches and uploaded into REDCap (Research Electronic Data Capture; REDCap Consortium). Participants received individualized survey links in recruitment emails and verified their school email accounts before proceeding. Once a participant with their unique student ID had completed the survey, they were no longer able to fill out another survey. Recruitment emails were staggered, with approximately 2700 students contacted per week, so that experiences throughout the spring and fall semesters would be captured. A total of 9 rounds of emails were sent in the spring 2021 semester, and 9 additional rounds were sent in the fall 2021 semester. Additionally, any spring 2021 nonresponders were given one additional opportunity to participate that fall. All questions were mandatory to answer before moving on to the next question, but every question had a "prefer not to answer" option. The survey was determined to be "complete" when the participant clicked the "submit" button at the end. In total, 2 automatic reminders were sent from REDCap if participants had provided consent and started the questionnaire but had not clicked the "submit" button. Once participants moved on to the next page of the survey, they were not able to go back and change an answer.

Surveys were comprised of multiple scales gathered from collaborating faculty across the university. After compilation and programming, the surveys were extensively tested for branching completeness and automation fluidity. Ongoing spot testing throughout the rollout process ensured that the recruitment and survey responses were imported correctly in REDCap.

Once participants completed the baseline survey, they entered their phone number to receive SMS text messages for the daily diary portion of the study. Daily diary prompts began on the day following baseline survey completion, sent as both an SMS text message and an email at 9 AM EST, with a reminder message 1 hour later to participants who had not yet completed that day's survey. To be considered valid, participants had to initiate the daily survey before midnight and complete the survey before 1:00 AM EST. The web-based consent and baseline survey were delivered securely via REDCap, which also coordinated all SMS text message and email survey invitations. For the daily diary portion of the study, REDCap invitations pushed out Qualtrics daily survey links.

A baseline survey was completed by 2068 young adult college students; 97.3% (n=2012) then participated in a 21-day diary study. All recruitment, consent, text messaging, reminders for the daily diary, and incentive payments were completed remotely.

### Ethical Considerations

This research was reviewed and approved by the Pennsylvania State University's institutional review board (approval number: STUDY00015710). Prior to completing the baseline survey, participants completed the informed consent process digitally. This contained all the consent elements: purpose of study, typical length of time to complete surveys, data storage and security, potential risks to participants, and contact information for the principal investigator and institutional review board. Participants had an opportunity to download and print the

consent form. Data from the baseline and daily surveys were linked using numeric identifiers; personal identifying information was removed prior to data analysis.

Participants were compensated US $15 for completing the baseline survey and US $2 per daily diary survey, with 2 possible high-compliance bonuses of US $5, one if participants completed at least 6 daily surveys during the first week and one if participants completed at least 12 daily surveys during the second and third weeks combined. Participants could earn up to US $67 with high study compliance. Additionally, participants received 1 raffle ticket for each high-compliance bonus, for a 1 in 100 chance of winning a US $50 gift card for high compliance in the first week and a 1 in 100 chance of winning a US $100 gift card for high compliance in the second and third weeks. Of the 2012 students who participated in the daily diary study, 1599 (79.5%) completed at least 6 of the first 7 days, and 1284 (63.8%) completed at least 12 of the last 14 days.

## Surveys

### *Baseline Survey*

Participants who completed the 15-minute baseline survey were automatically enrolled in a 21-day diary study. The baseline survey assessed a variety of topics, including demographics, college standing, health, COVID-19 prevention behaviors and beliefs, housing and food insecurity, antiracism behavior, experiences of bias or discrimination, and substance use.

### *Demographic, Individual, and Family Characteristics*

In the first part of the survey, general demographic questions were asked, including questions about participants' age, sex, gender, sexual or gender minority status, race or ethnicity, height, weight, current health conditions, and employment status during the academic year. Participants reported on their high school grade point average, recent and cumulative college grade point average, country of origin, political ideology, whether their permanent home was in or out of state, and their parents' highest degree (used to identify first-generation college students). Participants also reported on several different aspects of their relationship with their caregivers, including family cohesion and attachment, family conflict, identification with parents, and conflict behavior [8-11].

### *College-Specific Characteristics*

Participants reported their class standing and were asked about their group affiliations in college (honors program, Greek system, other philanthropic organizations), previous participation in the university's early start program, and current learning modes. During the spring survey, participants could participate in remote classroom lectures. The majority of classes were reinstated in person for the duration of the fall survey; however, a remote option was often available as well. Participants were also asked about their use of on-campus health services and their feelings of belonging at the university.

### *Food and Housing Insecurity*

Food insecurity was assessed by a validated 2-item screening instrument to assess participants' access to enough food every day. The screener asked about worries around food running out and not having enough money to buy more food [12]. Students' housing security, homelessness, and living situation were assessed with modules from the Researching Basic Needs in Higher Education instruments [13].

### *General Experiences of Discrimination and Antiracist Behavior*

Participants completed the 21-item Anti-racism Behavioral Inventory [14], which included items such as "When I hear people telling racist jokes and using negative racial stereotypes, I usually confront them." and "The police unfairly target Black men and Latinos." Participants also completed the 7-item Williams Everyday Discrimination Scale [15], which assessed perceptions such as people acting as if they think the individual is not smart or as if they are afraid of the individual.

### *Mental Health*

Depression symptoms were assessed using the 10-item Center for Epidemiologic Studies Depression Scale (CESD-R-10) [16]. Anxiety symptoms were assessed using the 6-item General Anxiety subscale of the Center for Collegiate Mental Health Instrument (CCAPS-34) [17]. Social anxiety was assessed using the 6-item Social Interaction Anxiety Scale (SIAS-6) [18]. The adult attention-deficit/hyperactivity disorder (ADHD) Self-Report Scale (ASRS, version 1.1) [19] was used to assess adult symptoms of ADHD. Participants also completed the 16-item Social Responsiveness Scale (SRS-2-S) [20] and an 8-item instrument to assess their level of flourishing [21].

### *Behavior*

Students' histories of alcohol, cannabis, nicotine (including vaping, cigarettes, and other tobacco products), and other illicit substance use behaviors were assessed, as was their perception of alcohol use on campus [22]. Recent sexual activity was assessed.

### *21-Day Daily Diary Survey*

Each morning, beginning the day after completing the baseline survey, participants were sent a link to that day's 5-minute survey. Participants were required to click on the link to initiate the daily survey by midnight and had until 1:00 AM to complete the survey. Daily diary surveys asked in-depth questions about substances used the day before (alcohol, marijuana, smoking, and other illicit substances) and their consequences, health behaviors, sleep, pleasurable experiences, and feelings of belonging, engagement, and well-being, as well as COVID-19 prevention behaviors. If participants did not report any substance use the day before, they were asked an alternative set of questions about food and housing insecurity and physical activity instead of the in-depth questions about substance use and consequences.

### *Substance Use, Consequences, and Context of Use*

Participants were asked if they used a substance (ie, alcohol, marijuana, cigarettes, prescription painkillers, heroin, psychedelics, sedatives, or stimulants). If they answered yes to any of the substances, they were asked follow-up questions regarding how much of each they took. If they answered yes to alcohol or marijuana, they were asked a follow-up question about impairment [23,24]. If a participant reported any substance use, they were asked about any positive (eg, feeling buzzed,

more energetic) or negative consequences (eg, blacking out or feeling nauseated) of using those substances [25]. For each substance participants reported using, they were asked where they used the substance (eg, home, a friend's house, and outdoors). They also reported if they used that substance around other people, and answered follow-up questions about who those people were.

### *Daily Experiences of Discrimination*

Discrimination was asked with 1 question, adapted from the Williams Daily Discrimination Scale, assessing how strongly they experienced bias or discrimination yesterday on a scale from 0=not at all to 5=very strongly [15].

### *Mental Health*

Daily stress was assessed on a 0-100 scale from 0=not at all to 100=very extremely, with follow-up questions asking if anything COVID-19–related increased that stress. Participants were then asked to rate how stressful school relationships with friends or roommates and family interactions were. Daily affect was assessed using a 10-item scale assessing different feelings and emotions. This experience sampling method affect scale, taken from Wichers et al [26], was modeled after the Positive and Negative Affect Schedule Instrument [27], specifically selecting high-loading items on the negative affect (ie, insecure, lonely, anxious, low, guilty, and suspicious) and positive affect (ie, cheerful, content, energetic, and enthusiastic) spectra. Participants also answered 2 questions assessing worry (ie, I worried what other people thought of me and I was worried that I would say or do the wrong things), drawn from the social anxiety scale from Kashdan and Steger [28].

### *Pleasure, Belonging, and Well-Being*

Experiences of pleasure were measured with 3 questions asking participants if they had any pleasure in activities yesterday and, if so, what type of activities they were and with what intensity they experienced pleasure [29]. Belonging questions asked the extent to which the participant felt like they belonged in the university environment [30]. Participants were asked 5 questions from the flourishing scale about whether they led a purposeful or meaningful life and were satisfied with their life, bored, or engaged [21].

### *Eating and Food Insecurity*

Participants reported where they ate yesterday. To assess food insecurity, they were asked whether they were worried about paying for food tomorrow and what their food situation was the day before (food insecurity screening instrument) [12]. To help balance the questionnaire length across all days, participants were asked these questions only on days when they reported no substance use.

### *Sleep and Physical Activity*

Each day, participants reported the time they went to bed, the time they woke up, and where they slept the night before (eg, college housing, apartment, and family home). Physical activity was assessed by 3 questions asking about vigorous and moderate-intensity exercise and sitting and reclining time. Participants were asked these questions only on days when they reported no substance use.

## Data Analysis

### *Scale Construction for Baseline Constructs*

For the *depression scale* (based on the CESD-10), 10 items with choices ranging from 0=rarely to 3=most of the time were summed to create a depression score, which ranged from 0 to 30. For the *anxiety scale* (based on CCAPS-34), 6 items with choices ranging from 0=not at all like me to 4=extremely like me were averaged to create an anxiety score, which ranged from 0 to 4. For the *social anxiety scale* (based on SIAS-6B), 6 items with choices ranging from 0 (not at all) to 4 (extremely) were summed to create a social anxiety score, which ranged from 0 to 24. The *ADHD symptoms scale* (based on ASRS_part A) was based on 6 items with response options from 0 (never) to 4 (very often); these items were converted to binary variables, wherein 3 of the items were scored as 1 if their value was greater than 1 (0 otherwise) and the remaining 3 items were scored as 1 if their value was greater than 2 (0 otherwise) [19,31]. The sum of these binary indicators created the ADHD symptoms scale, with a possible range of 0-6. For the *social responsiveness score* (based on the SRS-2-S), 16 items with response options from 0 (not true) to 3 (almost always true) were summed to create an SRS-2-S score, which ranged from 0 to 48. The original SRS-2 has 65 items, and scores could range from 0 to 195. The SRS-2-S scores were converted to an estimated SRS-2 score based on the recommended conversion table [32]. This estimated score could range from 5.7 (corresponding to SRS-2-S values of 0) to 181.5 (corresponding to SRS-2-S values of 48). For the *flourishing scale* at baseline, 8 items with response options ranging from 1=strongly disagree to 7=strongly agree were summed. The 8 items on this scale had high reliability (Cronbach $\alpha$=0.98). To calculate the *closeness with caregivers scale*, 4 items with response options ranging from 0 (almost never) to 100 (almost always) were averaged. The items asked about whether caregivers respected participants' feelings and encouraged them to talk about difficulties and whether they feel close to caregivers and tell caregivers about problems. This scale had high reliability (Cronbach $\alpha$=0.87).

### *Scale Construction for Daily Constructs*

For each observed day, the following scale scores were created. In total, 3 items tailored to assess aspects of flourishing in daily life were averaged to create a *daily flourishing scale*: All things considered, I am satisfied with my life yesterday; I was engaged and interested in my activities yesterday; and I led a purposeful and meaningful life yesterday. Each item had response options of 1=strongly disagree to 7=strongly agree. A *daily positive affect score* and a *daily negative affect score* were calculated using items from a 10-item scale where choices ranged from 1=never to 7=always [26]. In total, 4 of the items were averaged to create a positive affect score, and the remaining 6 items were averaged to create a negative affect score. A *daily worry score* was calculated by averaging 2 items that asked participants the extent to which they worried about what others thought of them and the extent to which they worried they would say or do the wrong things. Response options ranged from 1=not at all or slightly to 5=extremely. This scale had high reliability (Cronbach $\alpha$=0.88).

### *Statistical Modeling*

Descriptive statistics were calculated for a variety of constructs at baseline and in the daily study. Then, variance in daily response variables due to between- versus within-person sources was partitioned using multilevel models, with a separate intercept-only linear mixed model specified for each variable. In the case of a discrete outcome, this estimated percentage is an approximation; this is a reasonable approximation when the underlying probability of the response is not close to 0 or 1 [33]. All descriptive statistics and multilevel models were conducted using R (version 4.1.2; The R Foundation).

## Results

A summary of descriptive statistics from the baseline assessment appears in Table 1, and Table 2 presents descriptive statistics from the daily surveys, along with the proportion of variance that can be attributed to between-person versus person-level sources of variability.

**Table 1.** Descriptive statistics from the baseline survey.

| Baseline measure | Study sample (N=2068) | Population of UP[a] students (N=39,392) |
|---|---|---|
| Identify as female gender, n (%) | 1378 (66.6) | 18,639 (47.3) |
| **Race or ethnicity, n (%)** | | |
| Hispanic ethnicity | 162 (7.8) | 3029 (7.7) |
| Non-Hispanic Black | 56 (2.7) | 1713 (4.3) |
| Non-Hispanic White | 1433 (69.3) | 25,674 (65.2) |
| Non-Hispanic Asian | 307 (14.8) | 2651 (6.7) |
| Other and multiple race categories | 110 (5.3) | 6325 (16.1) |
| **Class standing, n (%)** | | |
| Lower undergraduates (first or second year) | 1039 (50.2) | 18,192 (46.2) |
| Upper undergraduates (third year or later) | 1025 (49.6) | 21,200 (53.8) |
| Foreign-born student, n (%) | 284 (13.7) | 3988 (10.1)[b] |
| First-generation student, n (%) | 348 (16.8) | 6621 (16.8) |
| In-state student status, n (%) | 1330 (64.3) | 23,289 (59.1) |
| Age (years), mean (SD) | 19.8 (1.3) | —[b] |
| **Employment status, n (%)** | | |
| Full-time | 20 (1.0) | — |
| Part-time | 771 (37.3) | — |
| Not employed | 1234 (59.7) | — |
| Sexual or gender minority, n (%) | 361 (17.5) | — |
| Engaged with Greek system, n (%) | 286 (13.8) | — |
| Food insecure (past month), n (%) | 395 (19.1) | — |
| Housing insecure (past month), n (%) | 170 (8.2) | — |
| Past-month alcohol use, n (%) | 1316 (63.6) | — |
| Past-month heavy episodic drinking, n (%) | 419 (20.3) | — |
| Past-month vaping, n (%) | 372 (18.0) | — |
| Past-month cigarette use, n (%) | 127 (6.1) | — |
| Past-month cannabis use, n (%) | 378 (18.3) | — |
| High school GPA[c], median (IQR) | 3.9 (0.4) | — |
| Cumulative college GPA, mean (SD) | 3.5 (0.4) | — |
| Antiracism behavior: scale of 21-105, mean (SD) | 67.7 (15.6) | — |
| General experiences of discrimination: scale of 1=almost daily to 6=never, mean (SD) | 4.85 (0.9) | — |
| Depression: scale of 0-30, mean (SD) | 11.0 (6.4) | — |
| General anxiety: scale of 0-4, mean (SD) | 1.40 (1.0) | — |
| Social anxiety: scale of 0-24, mean (SD) | 8.5 (6.0) | — |
| ADHD[d] symptoms: scale of 0-6, mean (SD) | 2.5 (1.8) | — |
| Social responsiveness: scale of 5.7-181.5, mean (SD) | 49.4 (25.6) | — |
| Flourishing: scale of 8-56, mean (SD) | 45.3 (7.4) | — |
| Closeness with caregivers: scale of 0-100, mean (SD) | 74.0 (20.7) | — |

[a]UP: University Park.

[b]Not applicable.

[c]GPA: grade point average.

[d]ADHD: attention-deficit/hyperactivity disorder.

**Table 2.** Descriptive statistics from daily diary study (based on 33,722 total person-days from n=2012 participants).

| Daily measure | Values | Proportion of variance, % | |
|---|---|---|---|
| | | Between-person | Within-person |
| Alcohol use day | 4902 (14.5)[a] | 19.4% | 80.6% |
| Heavy episodic drinking day | 2453 (7.3)[a] | 15.5% | 84.5% |
| Vaping day | 1888 (5.6)[a] | 69.8% | 30.2% |
| Cigarette use day | 254 (0.8)[a] | 40% | 60% |
| Cannabis use day | 1846 (5.5)[a] | 56% | 44% |
| Other illicit substance use day | 879 (2.6)[a] | 70.1% | 29.9% |
| Experienced interest or pleasure in activity | 24,771 (73.5)[a] | 40.7% | 59.3% |
| Stress: how stressful was yesterday on scale of 0=not at all to 100=extremely | 36.3 (27.2)[b] | 37.4% | 62.6% |
| How strongly experienced bias or discrimination: 0=not at all to 5=very strongly | 1.17 (0.5)[b] | 52.3% | 47.7% |
| Positive affect: 1=never to 7=always | 3.68 (1.5)[b] | 57.5% | 42.5% |
| Negative affect: 1=never to 7=always | 2.28 (1.1)[b] | 63.6% | 36.4% |
| Worried: 1=not at all or slightly to 5=extremely | 2.16 (1.1)[b] | 67.7% | 32.3% |
| Feelings of belonging at the university: 1=strongly disagree to 7=strongly agree | 5.1 (1.6)[b] | 70.7% | 29.3% |
| Flourishing: 1=strongly disagree to 7=strongly agree | 4.63 (1.5)[b] | 60.2% | 29.8% |
| Poor food situation in past 24 hours[c] | 6103 (23.5)[b] | 55.9% | 44.1% |
| Worried about how to pay for food tomorrow[c] | 855 (3.3)[b] | 52% | 48% |
| Slept in own dorm or room or apartment or family home last night[c] | 24,554 (94.4)[b] | 29.8% | 70.2% |
| Engaged in moderate or vigorous physical activity[c] | 22,095 (84.9)[b] | 44.8% | 55.2% |

[a]Values are n (%). Estimated percentages of variance are based on intraclass correlation coefficients in multilevel models; these numbers are approximations in the case of binary response variables.

[b]Values are mean (SD).

[c]Questions only asked on non–substance use days; thus, percentages are based on 26,020 days.

## Participant Characteristics at Baseline

A total of 2068 students completed the baseline survey, including 1378 (66.6%) female students. The participants' ages ranged from 18 to 24 (mean 19.8, SD 1.3) years at baseline. The sample included 1039 lower-level undergraduates (50.2% in the first or second year) and 1025 upper-level undergraduates (49.6% in the third year or later).

### *Sample Versus Population Characteristics*

Table 1 shows descriptive statistics for the data about the population of 39,392 undergraduate students. Data are from the Penn State data digest [34], compiled for the fall 2021 semester. The racial-ethnic composition of the sample (7.8% Hispanic or Latinx, 2.7% non-Hispanic Black, 69.3% non-Hispanic White, 14.8% non-Hispanic Asian, and 5.3% other or mixed races) was similar to that of the population of students at this campus (7.7% Hispanic or Latinx, 4.3% non-Hispanic Black, 65.2% non-Hispanic White, 6.7% non-Hispanic Asian, and 16.1% other or mixed races). The sample had a larger proportion of female students than the broader university population (47.3% in population), and slightly higher percentages of lower-level students (50.2% vs 46.2% in the population), foreign-born students (13.7% vs 10.1%), and in-state students (64.3% vs 59.1%). The proportion of first-generation students was 16.8% in both the sample and the population.

### *Additional Sample Characteristics*

Of participants who completed the baseline survey, 38.3% (n=791) were employed, 17.5% (n=361) identified as a sexual or gender minority, 13.8% (n=286) were affiliated with the Greek system, 19.1% (n=395) reported food insecurity, and 8.2% (n=170) reported housing insecurity. The prevalence of past-month substance use behaviors was 63.6% (n=1316) for any alcohol use, 20.3% (n=419) for heavy episodic drinking (ie, binge drinking), 18% (n=372) for vaping, 6.1% (n=127) for cigarette use, and 18.3% (n=378) for cannabis use. The median high school grade point average was 3.9 with an IQR of 0.4, and their mean cumulative grade point average in college was 3.5 (SD 0.4). Mean scale scores for antiracism behavior,

discrimination, depression, general anxiety, social anxiety, ADHD symptoms, social responsiveness, flourishing, and closeness with caregivers are provided in Table 1.

### Student Health and Well-Being in Daily Life

On 21 consecutive days, participants were prompted to complete a survey that assessed a comprehensive set of daily markers of health, health behaviors, and well-being. Students completed the daily survey in an average of 16.8 (SD 5.7) days.

#### *Substance Use, Stress, Discrimination, Affect, and Flourishing*

Of the 33,722 person-days, a minority were substance use days; 4902 days (14.5%) involved alcohol use, 2453 days with heavy episodic drinking (7.3%), 1888 days with vaping (5.6%), 254 days with cigarettes (0.8%), 1846 days with cannabis (5.5%), and 879 days with other illicit substance use (2.6%). Experiences of pleasure were reported on 24,771 of 33,722 days (73.5%). Mean scale scores for daily stress, experiences of bias or discrimination, positive affect, negative affect, worry, feelings of belonging, and flourishing are provided in Table 2.

#### *Additional Indicators of Well-Being*

Based on data from the 26,020 person-days with no reported substance use, a poor food situation was reported on 6103 days (23.5%); worry about how to pay for food tomorrow was reported on 855 (3.3%) days; students slept in their own dormitory, room, apartment, or family home on 24,554 (94.4%) days; and students engaged in moderate or vigorous physical activity on 22,095 (84.9%) days.

#### *Between- Versus Within-Person Variability*

Partitioning variance into between- versus within-person sources of variance for daily indicators of health and well-being provides insight into which constructs tend to be more person-specific and potentially explainable by person-level characteristics, as opposed to those that are more day-specific and potentially explainable by the contextual characteristics of a particular day. Between-person differences explained the majority of variance in daily vaping (69.8% attributable to between-person differences), cannabis use (56%), other illicit substance use (70.1%), experiences of bias or discrimination (52.3%), positive affect (57.5%), negative affect (63.6%), worry (67.7%), feelings of belonging at the university (70.7%), flourishing (60.2%), poor food situation (55.9%), and worry about how to pay for food (52%).

In contrast, within-person differences (ie, day-to-day fluctuations within individuals) explained the majority of the variance in daily alcohol use (80.6% attributable to within-person differences), heavy episodic drinking (84.5%), cigarette use (60%), experiences of pleasure (59.3%), stress (62.6%), where a student slept last night (70.2%), and physical activity (55.2%), indicating that day-level contextual factors may be more relevant for these daily experiences and behaviors.

## Discussion

### Principal Results

This study examined a wide range of health and well-being indicators among college students in daily life during the COVID-19 pandemic. Consistent with prior research on substance use in daily life, the proportion of variance attributed to within-person fluctuations was particularly high for daily alcohol use, heavy episodic drinking, and cigarette use, suggesting that these behaviors are highly sensitive to contextual characteristics [35-38]. In contrast, daily vaping behavior and other illicit substance use behavior tended to be most stable as person-level characteristics; this is consistent with recent research that has shown level of dependence to be positively associated with stability of behavior across days [39]. Variance in other aspects of daily mental health (belonging, flourishing, positive affect, negative affect, and worry) was also primarily attributable to between-person (ie, trait-like) differences. This finding is consistent with a coordinated analysis of 6 intensive longitudinal studies of positive and negative affect; variance decomposition of daily data reliably showed affect to be more strongly characterized by between-person differences [40].

These findings suggest where student programming and intervention efforts may be most effectively appropriated in this population. For example, available mental health services and interventions directed to certain individuals who would benefit most may be the best strategy for resource allocation on college campuses, whereas the provision of just-in-time mental health interventions on days of particularly high negative affect or worry or low positive affect may be less effective. In contrast, interventions designed to reduce problematic alcohol use and cigarette use may be best targeted toward risky contexts that students move in and out of throughout their days.

This type of intensive surveillance provides data that characterize not just population-average levels on aspects of health and well-being and their correlates, but also sheds light on day-to-day variability in student health and well-being and dynamic contexts that may place students at heightened risk for poor outcomes on a given day. By partitioning the variance of daily indicators of health and well-being, this study design provides the opportunity to determine which indicators of health and well-being are more trait-like and which are more state-like.

### Limitations and Future Work

One limitation of this study was the low rate of participation. This may produce nonresponse bias, which has been documented among college students, with survey respondents more likely to be female, socially engaged, and financially secure, with differences in personality compared to nonresponders (eg, [41]). Another limitation of this study was that rolling recruitment spanned from February through November 2020 (with no recruitment during the summer term); thus, students participated in the study at qualitatively distinct times during the COVID-19 pandemic. Most notably, vaccinations became widely available in mid-2020, which could have impacted student mental health and social behaviors. Thus, it may be important to determine whether these findings replicate in future analyses using daily diary data collected after the pandemic.

## Conclusions

Certain daily experiences and behaviors related to college students' health and well-being, such as vaping, negative affect, and feelings of belonging at one's university, distinguish at-risk individuals from other students. In contrast, other experiences and behaviors, such as alcohol use and cigarette use, stress, and physical activity, may be driven more by contextual risk factors operating on certain days. These findings suggest that some aspects of health and well-being may be most amenable to services that target students more generally, while others may be more amenable to interventions that target particularly risky students, days, or contexts.

Daily diary studies can inform practical policy recommendations to higher-education institutions, enabling them to make data-informed decisions on university closures and the development of programs to support student health and well-being. This approach is essentially a community-based surveillance system that can provide insight into student characteristics and contextual factors that place individuals at risk for negative outcomes.

## Acknowledgments

The data collection was funded internally by the Pennsylvania State University. The study was supported by awards (P50 DA039838) from the National Institute on Drug Abuse (NIDA). The NIDA did not have any role in study design, collection, analysis, and interpretation of the data; writing the report; and the decision to submit the report for publication. The work described here does not necessarily represent the official views of the National Institutes of Health. Our team acknowledges contributions to data collection made by Ms Jessica DeFelice, contributions to document preparation made by Ms Amanda Applegate, and comments on an early draft provided by Drs Maithreyi Gopalan and Danny Rahal.

## Data Availability

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

## Authors' Contributions

STL conceptualized the study, secured funding for the study, supervised the project, and produced the original draft and edited further versions of this manuscript. CW administered all aspects of data collection, compiled sources for all instruments used in the study, and reviewed and edited the manuscript. SB served as the data manager, conducted all analyses, and reviewed and edited the manuscript.

## Conflicts of Interest

None declared.

## References

1. Center for Collegiate Mental Health 2019 Annual Report: Global Health Education and Learning Incubator at Harvard University. 2020. URL: https://repository.gheli.harvard.edu/repository/collection/resource-pack-college-students-and-mental-health/resource/12372 [accessed 2024-07-12]
2. NA. Key substance use and mental health indicators in the United States: results from the 2019 national survey on drug use and health. Substance Abuse and Mental Health Services Administration 2020. [doi: 10.1037/e501902006-001]
3. White A, Hingson R. The burden of alcohol use: excessive alcohol consumption and related consequences among college students. Alcohol Res Curr Rev 2013;35(2):201-218. [doi: 10.1037/t69599-000]
4. Czeisler M, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
5. Wang X, Hegde S, Son C, Keller B, Smith A, Sasangohar F. Investigating mental health of US college students during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2020;22(9):e22817 [FREE Full text] [doi: 10.2196/22817] [Medline: 32897868]
6. Gopalan M, Linden-Carmichael A, Lanza S. College students' sense of belonging and mental health amidst the COVID-19 pandemic. J Adolesc Health 2022;70(2):228-233 [FREE Full text] [doi: 10.1016/j.jadohealth.2021.10.010] [Medline: 34893423]
7. Lanza ST, Whetzel CA, Linden-Carmichael AN, Newschaffer CJ. Change in college student health and well-being profiles as a function of the COVID-19 pandemic. PLoS One 2022;17(5):e0267724 [FREE Full text] [doi: 10.1371/journal.pone.0267724] [Medline: 35499988]
8. Bloom BL. A factor analysis of self-report measures of family functioning. Fam Process 1985;24(2):225-239. [doi: 10.1111/j.1545-5300.1985.00225.x] [Medline: 4018243]
9. Furstenberg FF, Cook TD, Eccles J, Elder GH. Managing to Make It: Urban Families and Adolescent Success. Chicago and London: University of Chicago Press; 2000.

10. Robin AL, Foster SL. Negotiating Parent-Adolescent Conflict: A Behavioral-Family Systems Approach. New York: Guilford Press; 2002.
11. Armsden GC, Greenberg MT. The inventory of parent and peer attachment: individual differences and their relationship to psychological well-being in adolescence. J Youth Adolesc 1987;16(5):427-454. [doi: 10.1007/BF02202939] [Medline: 24277469]
12. Hager ER, Quigg AM, Black MM, Coleman SM, Heeren T, Rose-Jacobs R, et al. Development and validity of a 2-item screen to identify families at risk for food insecurity. Pediatrics 2010;126(1):e26-e32. [doi: 10.1542/peds.2009-3146] [Medline: 20595453]
13. Crutchfield R, Maguire J. Researching Basic Needs in Higher Education: Qualitative and Quantitative Instruments to Explore a Holistic Understanding of Food and Housing Insecurity. California: California State University; 2017:1-80.
14. Pieterse AL, Utsey SO, Miller MJ. Development and initial validation of the anti-racism behavioral inventory (ARBI). Couns Psychol Q 2015;29(4):356-381. [doi: 10.1080/09515070.2015.1101534]
15. Williams DR, Yu Y, Jackson JS, Anderson NB. Racial differences in physical and mental health: socio-economic status, stress and discrimination. J Health Psychol 1997;2(3):335-351 [FREE Full text] [doi: 10.1177/135910539700200305] [Medline: 22013026]
16. Zhang W, O'Brien N, Forrest JI, Salters KA, Patterson TL, Montaner JSG, et al. Validating a shortened depression scale (10 item CES-D) among HIV-positive people in British Columbia, Canada. PLoS One 2012;7(7):e40793 [FREE Full text] [doi: 10.1371/journal.pone.0040793] [Medline: 22829885]
17. Locke BD, McAleavey AA, Zhao Y, Lei PW, Hayes JA, Castonguay LG, et al. Development and initial validation of the counseling center assessment of psychological symptoms–34. Meas Eval Couns Dev 2017;45(3):151-169. [doi: 10.1177/0748175611432642]
18. Peters L, Sunderland M, Andrews G, Rapee RM, Mattick RP. Development of a short form social interaction anxiety (SIAS) and social phobia scale (SPS) using nonparametric item response theory: the SIAS-6 and the SPS-6. Psychol Assess 2012;24(1):66-76. [doi: 10.1037/a0024544] [Medline: 21744971]
19. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, et al. The world health organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. Psychol Med 2005;35(2):245-256. [doi: 10.1017/s0033291704002892] [Medline: 15841682]
20. Constantino JN, Gruber CP. Social Responsiveness Scale: SRS-2. Torrance, CA: Western Psychological Services; 2012.
21. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, et al. New well-being measures: short scales to assess flourishing and positive and negative feelings. Soc Indic Res 2010;97(2):143-156. [doi: 10.1007/s11205-009-9493-y]
22. Graupensperger S, Turrisi R, Jones D, Evans MB. Longitudinal associations between perceptions of peer group drinking norms and students' alcohol use frequency within college sport teams. Alcohol Clin Exp Res 2020;44(2):541-552 [FREE Full text] [doi: 10.1111/acer.14270] [Medline: 31943243]
23. Cloutier RM, Calhoun BH, Lanza ST, Linden-Carmichael AN. Assessing subjective cannabis effects in daily life with contemporary young adult language. Drug Alcohol Depend 2022;230:109205 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.109205] [Medline: 34890928]
24. Linden-Carmichael AN, Calhoun BH. Measuring subjective alcohol effects in daily life using contemporary young adult language. Exp Clin Psychopharmacol 2022;30(2):151-158 [FREE Full text] [doi: 10.1037/pha0000447] [Medline: 33591776]
25. Lee CM, Cronce JM, Baldwin SA, Fairlie AM, Atkins DC, Patrick ME, et al. Psychometric analysis and validity of the daily alcohol-related consequences and evaluations measure for young adults. Psychol Assess 2017;29(3):253-263 [FREE Full text] [doi: 10.1037/pas0000320] [Medline: 27196690]
26. Wichers M, Peeters F, Rutten BPF, Jacobs N, Derom C, Thiery E, et al. A time-lagged momentary assessment study on daily life physical activity and affect. Health Psychol 2012;31(2):135-144. [doi: 10.1037/a0025688] [Medline: 21988094]
27. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
28. Kashdan TB, Steger MF. Expanding the topography of social anxiety. An experience-sampling assessment of positive emotions, positive events, and emotion suppression. Psychol Sci 2006;17(2):120-128. [doi: 10.1111/j.1467-9280.2006.01674.x] [Medline: 16466419]
29. Heininga VE, van Roekel E, Ahles JJ, Oldehinkel AJ, Mezulis AH. Positive affective functioning in anhedonic individuals' daily life: anything but flat and blunted. J Affect Disord 2017;218:437-445 [FREE Full text] [doi: 10.1016/j.jad.2017.04.029] [Medline: 28531841]
30. Walton GM, Cohen GL. A question of belonging: race, social fit, and achievement. J Pers Soc Psychol 2007;92(1):82-96. [doi: 10.1037/0022-3514.92.1.82] [Medline: 17201544]
31. Adler L, Kessler R, Spencer T. Adult ADHD Self-Report Scale-v1. 1 (ASRS-v1. 1) Symptom Checklist. New York, NY: World Health Organization; 2003.
32. Lyall K, Hosseini M, Ladd-Acosta C, Ning X, Catellier D, Constantino JN, program collaborators for Environmental influences on Child Health Outcomes. Distributional properties and criterion validity of a shortened version of the social responsiveness scale: results from the ECHO program and implications for social communication research. J Autism Dev Disord 2021;51(7):2241-2253 [FREE Full text] [doi: 10.1007/s10803-020-04667-1] [Medline: 32944847]

33. Goldstein H, Browne W, Rasbash J. Partitioning variation in multilevel models. Understanding statistics: statistical issues in psychology, education, and the social sciences. Sciences 2002;1(4):223-231. [doi: 10.1207/s15328031us0104_02]
34. Penn State Data Digest. 2021. URL: https://datadigest.psu.edu/ [accessed 2024-07-12]
35. Maggs JL, Williams LR, Lee CM. Ups and downs of alcohol use among first-year college students: number of drinks, heavy drinking, and stumble and pass out drinking days. Addict Behav 2011;36(3):197-202 [FREE Full text] [doi: 10.1016/j.addbeh.2010.10.005] [Medline: 21106298]
36. Colder CR, Lloyd-Richardson EE, Flaherty BP, Hedeker D, Segawa E, Flay BR, Tobacco Etiology Research Network. The natural history of college smoking: trajectories of daily smoking during the freshman year. Addict Behav 2006 Dec;31(12):2212-2222. [doi: 10.1016/j.addbeh.2006.02.011] [Medline: 16616816]
37. Park CL, Armeli S, Tennen H. The daily stress and coping process and alcohol use among college students. J Stud Alcohol 2004;65(1):126-135. [doi: 10.15288/jsa.2004.65.126] [Medline: 15000512]
38. Piasecki TM, Jahng S, Wood PK, Robertson BM, Epler AJ, Cronk NJ, et al. The subjective effects of alcohol-tobacco co-use: an ecological momentary assessment investigation. J Abnorm Psychol 2011;120(3):557-571 [FREE Full text] [doi: 10.1037/a0023033] [Medline: 21443289]
39. Buu A, Cai Z, Li R, Wong SW, Lin HC, Su WC, et al. Validating e-cigarette dependence scales based on dynamic patterns of vaping behaviors. Nicotine Tob Res 2021;23(9):1484-1489 [FREE Full text] [doi: 10.1093/ntr/ntab050] [Medline: 33758949]
40. Scott SB, Sliwinski MJ, Zawadzki M, Stawski RS, Kim J, Marcusson-Clavertz D, et al. A coordinated analysis of variance in affect in daily life. Assessment 2020;27(8):1683-1698 [FREE Full text] [doi: 10.1177/1073191118799460] [Medline: 30198310]
41. Porter SR, Whitcomb ME. Non-response in student surveys: the role of demographics, engagement and personality. Res High Educ 2005;46(2):127-152. [doi: 10.1007/s11162-004-1597-2]


## Abbreviations

**ADHD:** attention-deficit/hyperactivity disorder
**ASRS:** Adult ADHD Self-Report Scale
**CCAPS-34:** Center for Collegiate Mental Health Instrument
**CESD-R-10:** Center for Epidemiologic Studies Depression Scale
**REDCap:** Research Electronic Data Capture
**SIAS-6:** Social Interaction Anxiety Scale
**SRS-2-S:** Social Responsiveness Scale


*Edited by T de Azevedo Cardoso; submitted 15.01.23; peer-reviewed by J Abbas, L Yang; comments to author 02.09.23; revised version received 22.01.24; accepted 28.06.24; published 23.08.24.*

*Please cite as:*
*Lanza ST, Whetzel C, Bhandari S*
*Health and Well-Being Among College Students in the United States During the COVID-19 Pandemic: Daily Diary Study*
*Interact J Med Res 2024;13:e45689*
*URL: https://www.i-jmr.org/2024/1/e45689*
*doi:10.2196/45689*
*PMID:39178037*

Original Paper

# Unlocking the Potential of Secondary Data for Public Health Research: Retrospective Study With a Novel Clinical Platform

Christopher Gundler[1*], MSc; Karl Gottfried[1*]; Alexander Johannes Wiederhold[1], MD; Maximilian Ataian[1]; Marcus Wurlitzer[2], Dr rer nat; Jan Erik Gewehr[2], Dr rer nat; Frank Ückert[1], Dr med

[1]Institute for Applied Medical Informatics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Research Data Facility, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[*]these authors contributed equally

**Corresponding Author:**
Christopher Gundler, MSc
Institute for Applied Medical Informatics
University Medical Center Hamburg-Eppendorf
Martinistrasse 52
Hamburg, 20246
Germany
Phone: 49 40741054979
Email: c.gundler@uke.de

## Abstract

**Background:** Clinical routine data derived from university hospitals hold immense value for health-related research on large cohorts. However, using secondary data for hypothesis testing necessitates adherence to scientific, legal (such as the General Data Protection Regulation, federal and state protection legislations), technical, and administrative requirements. This process is intricate, time-consuming, and susceptible to errors.

**Objective:** This study aims to develop a platform that enables clinicians to use current real-world data for testing research and evaluate advantages and limitations at a large university medical center (542,944 patients in 2022).

**Methods:** We identified requirements from clinical practitioners, conceptualized and implemented a platform based on the existing components, and assessed its applicability in clinical reality quantitatively and qualitatively.

**Results:** The proposed platform was established at the University Medical Center Hamburg-Eppendorf and made 639 forms encompassing 10,629 data elements accessible to all resident scientists and clinicians. Every day, the number of patients rises, and parts of their electronic health records are made accessible through the platform. Qualitatively, we were able to conduct a retrospective analysis of Parkinson disease over 777 patients, where we provide additional evidence for a significantly higher proportion of action tremors in patients with rest tremors (340/777, 43.8%) compared with those without rest tremors (255/777, 32.8%), as determined by a chi-square test ($P$<.001). Quantitatively, our findings demonstrate increased user engagement within the last 90 days, underscoring clinicians' increasing adoption of the platform in their regular research activities. Notably, the platform facilitated the retrieval of clinical data from 600,000 patients, emphasizing its substantial added value.

**Conclusions:** This study demonstrates the feasibility of simplifying the use of clinical data to enhance exploration and sustainability in scientific research. The proposed platform emerges as a potential technological and legal framework for other medical centers, providing them with the means to unlock untapped potential within their routine data.

***(Interact J Med Res 2024;13:e51563)*** doi:10.2196/51563

**KEYWORDS**

secondary use; hypothesis testing; research platform; clinical data; Parkinson disease; data; health-related research; health data; electronic health record; EHR; tremor

## Introduction

In recent years, there has been a growing interest in using clinical routine data, especially electronic medical records, for research [1]. Known as secondary data use, this practice is significantly influenced by legislative actions such as the Health Information Technology for Economic and Clinical Health Act in the United States of America and publicly founded initiatives

like the Medical Informatics Initiative (MII) in Germany [2,3]. University hospitals now serve as central hubs in bridging the gap between research and patient care. Achieving connectivity between previously isolated data silos necessitates adherence to detailed standards, such as the custom FHIR (Fast Healthcare Interoperability Resources) profile "Kerndatensatz" in Germany and effective communication among diverse stakeholders within national health systems [4]. Despite the inherent complexity, these concerted efforts are expected to establish a new foundation for evidence-based science.

Despite the advantages of adopting a state-wide approach to secondary data use, we contend that certain research could be more effectively conducted at the local level within individual hospitals. Specifically, we have identified 2 critical use cases where hypothesis testing on unmapped raw data is essential for advancing evidence-based medicine:

First, to verify hypotheses derived from clinical practice on a larger database, clinicians should be able to validate their experiences easily through data-driven investigations. These studies may involve data elements not comprehensively covered by standardized data sets. In addition, expecting clinicians to create intricate mappings between used data elements and state-wide standards may prove ineffective.

Second, the replication of existing publications to assess their generalizability must always consider the local context. Conducting public health research, for example on large cohorts of patients with Parkinson disease (PD), might miss important external factors [5,6]. Accordingly, testing external validity on a schema applied in practice rather than one developed for collaboration appears more appropriate.

Beyond the clinical perspective, local solutions may better accommodate regional or local legal requirements, as many collaborative standardizations tend to converge on the "smallest common divisor" between partners. Consequently, the implementation of complementary systems for secondary data analysis at both the local and global levels is deemed appropriate.

Developments in recent years have led to the emergence of research platforms that enable the analysis of clinical data in compliance with data protection guidelines. Notable examples include EPOCH and ePRISM (IP-ITT Corporation) [7], KETOS (Friedrich-Alexander University Erlangen-Nürnberg) [8], and Medical-Blocks (University of Bern) [9]. These platforms provide environments that allow clinical scientists to train and deploy statistical models. However, their primary focus is on translating these models back into clinical practice rather than testing hypotheses through the secondary use of data. In addition, works such as EHR4CR (Electronic Health Records for Clinical Research) [10] have implemented infrastructures that enable the use of clinical data across multiple European sites in a secure and privacy-preserving manner without focusing on the subsequent analysis. Given our knowledge, a platform for hypothesis testing on routine data has not been implemented and evaluated in clinical reality.

This paper addresses this gap in research by introducing and evaluating a novel platform explicitly designed for hypothesis testing on clinical routine data. Starting by collecting the requirements of clinicians, we strive to design and implement a modular and, consequently, reusable platform. Similar to other states, the federal law of Hamburg permits pseudonymized retrospective data analysis without patient consent given specific guarantees regarding data protection. The platform ensures those guarantees in accordance with all European and German laws and is directly integrated into the technical infrastructure of the University Medical Center Hamburg-Eppendorf (UKE). The evaluation process encompasses quantitative assessment, exemplified by the replication of a public health finding in the context of PD, and qualitative evaluation through the examination of clinicians' use in real-world clinical scenarios. This dual-pronged evaluation strategy aims to judge both the quantitative efficacy and practical use of the proposed platform in clinical reality.

## Methods

### Technical Considerations

For developing the platform, we examined the challenges of using routine hospital data for hypothesis testing through extensive communication with the different business divisions of the UKE, like the infrastructure department, division for information technology, research data facility, data protection officers, and internal boards. Informed by the project meetings and discussions with clinicians as later users, we identified and prioritized 4 critical process components necessitating optimization.

#### *Defining Appropriate Hypotheses*

Precise hypothesis formulation relies on a thorough understanding of metadata within the clinical information system. For researchers, the accessibility of relevant data fields may not be immediately evident. Challenges arise from both nontechnical limitations and the opacity of data type and structure. Filtering cohorts based on specific criteria may yield statistically inappropriate sizes, and requested data may be inadequately recorded [11]. The feasibility of research ideas is thus not guaranteed, necessitating extensive consultation with data integration experts for hypothesis refinement.

#### *Obtaining Data From the Infrastructure*

Efficient storage and retrieval of routine hospital data are crucial for medical treatment and research. Hospitals use diverse IT architectures, often a mix of specialized systems with proprietary data structures and nonstandardized file formats. Access and control vary widely, from centralized systems to more federated approaches led by individual clinics. Clinicians aiming to test hypotheses face challenges in accessing required documentation, understanding these structures, and communicating with the responsible data manager.

#### *Analysis of the Hypotheses*

To facilitate hypothesis tests, clinicians expressed a need for a comprehensive and heterogeneous array of tools, encompassing table-based software and standard scripting languages like Python (Python Software Foundation) or R (R Foundation for Statistical Computing). Established research data management

platforms, such as Kaggle (Google) [12], Paperspace Gradient (DigitalOcean) [13], Colab (Google) [14], or CodaLab (Microsoft Research) [15], provide ideal support for efficient data analysis: An integrated and simplified development environment, a separate space for data analysis with access to high-performance computing, and the ability to communicate and collaborate with other users of the research community. Rather than developing a novel solution, leveraging a platform that accommodates diverse analysis methods appears to be a pragmatic approach.

### *Reuse of Established Components*

Based upon the preliminary work of the MII and the existing research landscape, the following tools were explored as relevant in the context of our work.

### *Data Integration*

Data integration centers (DICs) enable the cross-site and cross-institutional use of digital health data from patient care and biomedical research in Germany [2,3,16]. All DICs are located at university medical sites and have access to routinely collected patient data. To this end, they build up interoperable databases with quality-assured and internationally harmonized data (based on HL7 [Health Level Seven International] FHIR) and metadata. These are made available in anonymized form through trustees. DICs make an important contribution to the development of a research-orientated infrastructure for the German health care system. The first use cases using the functionalities are already in operation [17]. These functionalities are reusable and valuable for our work. For further details, we refer to the literature regarding the MII [2,3].

### *Data Usage Considerations*

European, national, and local laws govern the use of sensitive routine data. Those projects necessitate ethical approval and explicit consent, a crucial yet burdensome process for both researchers and ethics committee members. As the legislators have already identified the need for simplification, we were able to use §12 of the "Hamburgisches Landeskrankenhausgesetz" [18]. This statute permits pseudonymized retrospective data analysis without patient consent, allowing us to forego consent-based data usage. With approval from the ethics committee for hypothesis testing, the board and the individual researcher might focus more on the research question rather than time-consuming bureaucratic processes. Without this general approval for hypothesis testing, researchers would normally not be able to query the data without extensive knowledge regarding the infrastructure and the law. Furthermore, the UKE has established an independent trust center, which is largely autonomous in legal terms. This center uses suitable pseudonymization techniques to safeguard patient data identity.

### *Metadata Processing*

The processing of metadata is crucial in the context of data harmonization with multiple data sources, as intended in this project. Metadata repositories (MDRs) enable the structuring of data for the technical extract, transform, and load (ETL) process. They are also applications that make the syntax and semantics of the data understandable for the end user. Both attributes are relevant in our context. Numerous systems have already been tested and evaluated in use [15-17]. In this case, we prefer an MDR that is a further development of the already used Samply.MDR [19], an ISO/IEC 11179-based metadata repository built on a graph-based backend, making the MDR applicable to many hierarchical data structures.

## Quantitative and Qualitative Analysis

In the evaluation of the proposed tool, a dual assessment was conducted, encompassing a qualitative analysis of its suitability for replicating a public health-focused study and a quantitative examination of clinicians' usage behavior within the hospital.

### *Qualitative Analysis: Hypothesis Testing*

The capabilities of the proposed platform for testing scientific hypotheses appear to be valuable for replicating studies in other cohorts. Comparable to existing publications [8], we applied the platform to underscore its efficacy in promoting sound scientific practices and for examining the generalizability of findings regarding the circumstances present at a specific hospital.

Due to its notable clinical implications [20,21], we chose PD as a neurodegenerative disorder of interest for which routine data may provide helpful insights. The International Parkinson and Movement Disorder Society (MDS) developed a scoring system to measure the severity of PD motor symptoms. This movement test is called the Unified Parkinson's Disease Rating Scale (UPDRS) and is widely used in clinical routine [22]. While postural, kinetic, and isometric tremor are subcategories of action tremor, the isometric tremor is difficult to measure in routine clinical settings and is not routinely assessed [23,24]. Nonetheless, the exact relationship between these distinct types of tremors remains incompletely understood.

Motivated by the findings of Gupta et al [25], our objective is to validate their proposed correlation between rest tremor and action tremor in patients with PD [26]. Consequently, we aim to replicate their observation of a significantly higher prevalence of action tremor in individuals also experiencing resting tremor.

By leveraging the proposed platform, we gained access to routine data, expanding beyond the use of public data sets used in the original study: The Parkinson Progression Marker Initiative (PPMI) [27], the Fox Investigation for New Discovery of Biomarkers (BioFIND) [28], and the Parkinson's Disease Biomarkers Program (PDBP) [29] data sets are 3 distinct clinical oriented, observational studies collecting relevant disease-specific data from patients with PD. The PPMI study focuses on early-stage patients with PD who have recently been diagnosed and are not yet receiving dopaminergic treatment. In contrast, the BioFIND and PDBP studies encompass patients at varying stages of PD, ranging from moderate to advanced and early to advanced, respectively. Consequently, the PPMI data set exclusively includes patients in the medication-off state, while the latter 2 data sets include patients in both the medication-off and medication-on states.

As the first step of the analysis, we identified those forms within the clinical information system used to store classifications according to the MDS-UPRDS. The platform facilitated the

selection of a well-defined cohort, ensuring precise inclusion criteria for the data query. Accordingly, we included all patients with the designated *ICD-10-GM* (*International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, German Modification*) code for Parkinson disease (G20). Furthermore, we limited our cohort to admissions that occurred between February 24, 2018, and February 24, 2023. Although age limitations can be applied, we did not impose any restrictions for the presented cohort. Leveraging the aforementioned criteria, we executed the query and retrieved all corresponding records stored in the system.

### *Quantitative Analysis*

For the quantitative analysis, we focus on performance indicators critical for assessing the relevance of our platform in clinical reality. The practical use of the platform is measured with the cumulative probability distribution and the absolute number of requests after the initialization of the platform. The waiting times are critical for user experience, which is expected of the researcher's experience when they receive the requested data.

### Ethical Considerations

Based on the proposed pipeline for pseudonymization and data security, the ethics committee of the Hamburg chamber of physicians agreed on approval for all hypothesis tests conducted through the platform (2022-100891-BO-ff).

## Results

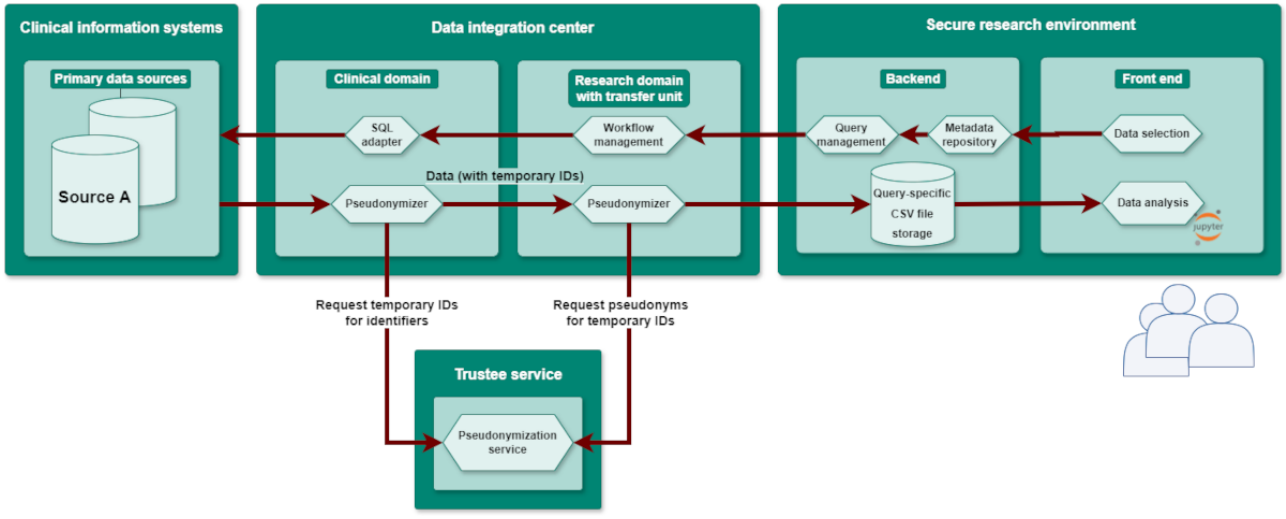
### Technical Realization Within Clinical Reality

The central implementation detail of the platform is the usage of the established systems through strategic interfacing within the DIC. Notably, it circumvents the integration phase by directly querying the databases of clinical systems. The platform's backend is incorporated into the DIC's clinical and research domains, using the central trustee service for pseudonymization and the transfer unit for managing workflows and facilitating data delivery to researchers. The underlying architecture is constructed using standardized web technologies, specifically HTTPS and REST (Representational State Transfer). These procedural steps are fully automated, furnishing transparent feedback on the ongoing progress (Figure 1).

The data architecture is organized into 3 primary sections: clinical information systems, data integration center, and secure research environment. A researcher initiates hypothesis testing using the web application in the front end of the secure research environment. Data selection for this process is facilitated through a catalog of data elements supplied by the metadata repository. The data integration center, which is divided into 2 domains, the clinical domain, and the research domain, is responsible for converting the researcher's query into potentially multiple SQL (Structured Query Language) commands to retrieve data from the clinical information system's database. The output is exported as a CSV (comma-separated values) file and subjected to the pseudonymization triangle, where identifiers like medical record numbers and visit numbers are replaced with temporary IDs by the clinical domain before transmitting the data set to the research domain. The trustee service ensures that data from various sources are assigned consistent pseudonyms, maintaining the integrity of the hypothesis testing context. Subsequently, the pseudonymized data is stored in the query-specific CSV file storage within the backend of the secure research environment. Researchers have access to these data sets for further analysis and can use analytical tools such as Jupyter notebooks, which are available on the front end.

Text data or images that contain sensitive information are not exported. Subsequently, the research domain decodes temporary IDs into definitive pseudonyms and stores the data set for subsequent researcher access. Both the clinical and research domains obtain temporary IDs and pseudonyms from the trustee service, configured to issue unique ones for each identifier type (eg, medical record number or visit number). The linkage between original IDs and pseudonyms remains confidential to both domains; solely, the trustee service retains this information throughout the project's duration. After this process, the resulting file is automatically downloaded into a separate network designed for research and stored on a file system accessible only to the requesting clinician.

**Figure 1.** Components and data processing architecture of the platform. CSV: comma-separated value; SQL: Structured Query Language.



Using a widely recognized solution for both clinicians and data scientists, we incorporated JupyterLab, a prominent open-source web-based development environment, as the principal front end for ensuing data analyses. As a result, the proposed platform empowers users to leverage a diverse array of tools and libraries on the 639 forms encompassing 10,629 data elements that we have made accessible from clinical routine.

### Qualitative Results: Example of a Hypothesis Testing

For the qualitative assessment, all patients with the designated *ICD-10-GM* code G20 and admissions that occurred between February 24, 2018, and February 24, 2023, were included. The majority of patients underwent multiple assessments during their hospital visit. We only considered the first assessment to ensure independent samples and discarded subsequent assessments. This was necessary since we were mainly interested in the patient cohort with any of the 3 basic tremor types rather than the overall occurrence of all tremors assessed at any given time point. The decision to choose the first assessment was made since not every subject was assessed more than once, but always at the beginning of their hospitalization, thereby ensuring uniformity in tremor severity assessments shortly after admission. Afterward, we derived the subtypes described in the original work by considering the MDS-UPDRS items 3.17, 3.15, and 3.16 as surrogates for rest tremor, postural tremor of the hands, and kinetic tremor of the hands, respectively. As a result, we were able to include 777 patients in our qualitative assessment.

Table 1 presents the prevalence of the primary tremor types and the association between rest tremor and action tremor. The table includes 4 distinct data sets, with the first 3 data sets obtained from Gupta et al [25] and the fourth data set corresponding to our analysis conducted using the proposed tool (UKE). The provided values for rest tremor, postural tremor, and kinetic tremor represent the count of patients with PD exhibiting each respective tremor type while at rest, while holding their hands stretched out, or during a finger-to-nose maneuver, respectively. The severity rating for each tremor type is equal to or above 1, as outlined in the MDS-UPDRS guideline. The aggregated values presented in the table were derived following the published protocol.

**Table 1.** Comparison of the tremor subtypes and their occurrences within the cohorts reported by Gupta et al [25]. In addition, the last column shows the obtained results from routine data based on the proposed platform.

| | PPMI[a] (N=423), n (%) | BioFIND[b] (N=118), n (%) | PDBP[c] (N=874), n (%) | UKE[d] (N=777), n (%) |
|---|---|---|---|---|
| Rest tremor | 290 (68.6) | 75 (63.6) | 459 (52) | 340 (43.8) |
| Pure rest tremor | 87 (20.6) | 15 (12.7) | 104 (11.8) | 57 (7.3) |
| Action tremor | 156 (36.9) | 46 (39) | 316 (35.8) | 255 (32.8) |
| Pure action tremor | 40 (9.5) | 10 (8.5) | 87 (9.9) | 76 (9.8) |
| Postural tremor | 223 (52.7) | 69 (58.5) | 412 (46.7) | 416 (53.5) |
| Pure postural tremor | 18 (4.3) | 8 (6.8) | 31 (3.5) | 72 (9.3) |
| Kinetic tremor | 217 (51.3) | 61 (51.7) | 463 (52.5) | 301 (38.7) |
| Pure kinetic tremor | 23 (5.4) | 6 (5.1) | 86 (9.8) | 31 (4) |
| No tremor | 52 (12.3) | 19 (16.1) | 211 (23.9) | 258 (33.2) |
| Any tremor | 317 (87.7) | 99 (83.9) | 663 (75.2) | 519 (66.8) |
| All tremor | 116 (27.4) | 36 (30.5) | 229 (26) | 179 (23) |

[a]PPMI: Parkinson Progression Marker Initiative.

[b]BioFIND: Fox Investigation for New Discovery of Biomarkers.

[c]PDBP: Parkinson's Disease Biomarkers Program.

[d]UKE: University Medical Center Hamburg-Eppendorf.

Our results represent a cohort of patients with PD irrespective of any dopaminergic treatment since many patients lack information regarding medication status due to the subsequent addition of this data field into the clinical information system. Through our data analysis, we observed a prevalence of 43.8% (340/777) for rest tremors and 7.3% (57/777) for pure rest tremors within the cohort. In contrast, the prevalence of total action tremor was 32.8% (255/777), with a corresponding occurrence of 9.8% (76/777) for pure action tremor. The incidences of postural tremor and pure postural tremor were found to be 53.5% (416/777) and 9.3% (72/777), respectively. We identified a prevalence of 38.7% (301/777) for kinetic tremor and 4.0% (31/777) for pure kinetic tremor. Finally, we calculated the occurrence of patients exhibiting all 3 tremor types simultaneously, the absence of any tremor, and the presence of at least 1 tremor type, resulting in proportions of 23.0% (179/777), 33.2% (258/777), and 66.8% (519/777), respectively. These relative figures closely resemble the reported values from the original authors. Importantly, we also observed a significantly higher proportion of action tremors in patients with rest tremors (43.8%) compared with those without rest tremors (32.8%), as determined by a chi-square test ($P$<.001).

Our analysis of routine data has yielded additional evidence that aligns with the published findings, suggesting that action tremor may be part of a broader tremor syndrome observed in PD. This discovery emphasizes the need for a more dynamic approach to tremor classification, considering the progressive

worsening of rest tremor severity over time [30] and its potential association with the occurrence of action tremor. Specifically, our data set corroborates the previous findings by Gupta et al [25], which propose a relationship between rest tremor and the emergence of action tremor. The data we have obtained further suggests that action tremor may represent a manifestation of the underlying basal ganglia disease, highlighting the potential requirement for additional neuroimaging studies to elucidate this connection.

## Quantitative Results

In the realm of quantitative results, we focus on performance indicators critical for assessing the relevance of our platform within the clinical reality. To that end, we compiled a comprehensive list of successful queries executed using our proposed tool before October 30, 2023. Subsequently, we exclude queries carried out by members of the development teams, as they were primarily intended for debugging purposes.

Figure 2 illustrates the cumulative probability distribution of all incorporated queries across the temporal dimension. A conspicuous observation is the initial absence of queries in the early phase, signifying a notable delay in the adoption of the tool by clinical researchers, spanning nearly 6 months. The discernible rightward shift indicates an increasing interest among researchers following an initial habituation period. Nevertheless, the following data points reveal a marked acceleration in query use, with over 50% of the total queries executed within the most recent 3-month period.

Our platform offers the unique advantage of accessing multiple systems integral to clinical care. However, it's important to note that these platforms are not optimized for the specific nature of the queries in question. Substantial delays in data retrieval could significantly impede the quality of research conducted using our tool. Consequently, we analyzed to evaluate the waiting times experienced by clinical researchers before they received the requested data.

Figure 3 displays the distribution of the time it took for the queried data to become available to the researcher. The plot reveals a notable range of waiting times. While more than 50% of all requests were processed within a time frame of 50 hours, the longest queries extended to nearly a week. The pronounced initial ascent up to the median highlights the prompt reception of a substantial proportion of data despite the existence of instances where requests experience prolonged processing durations.

**Figure 2.** Cumulative probability distribution of researcher-initiated queries over time, starting from the public announcement of the platform, as extracted from the platform logs.

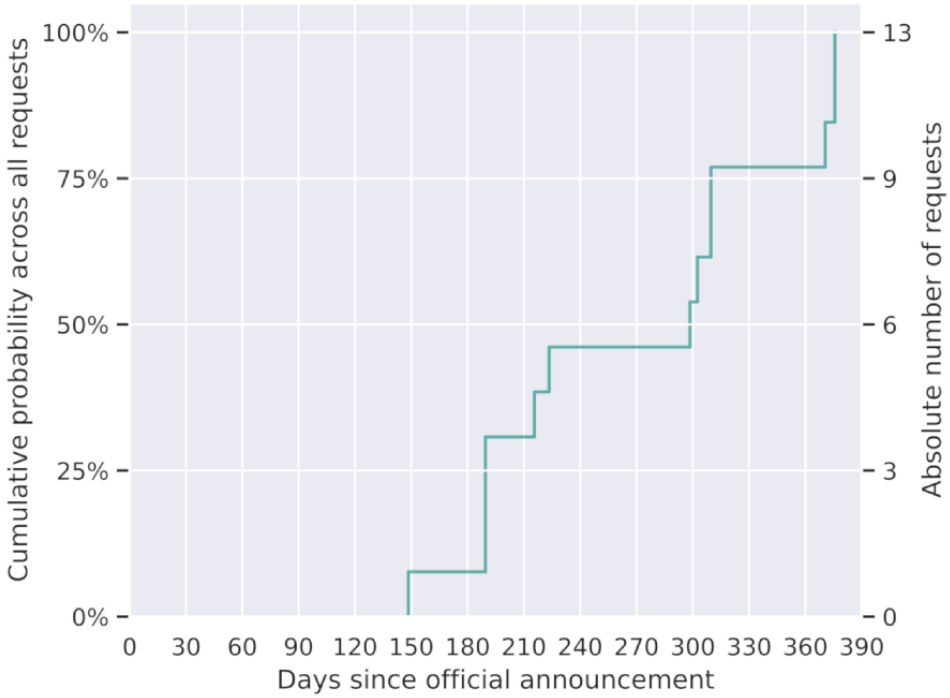
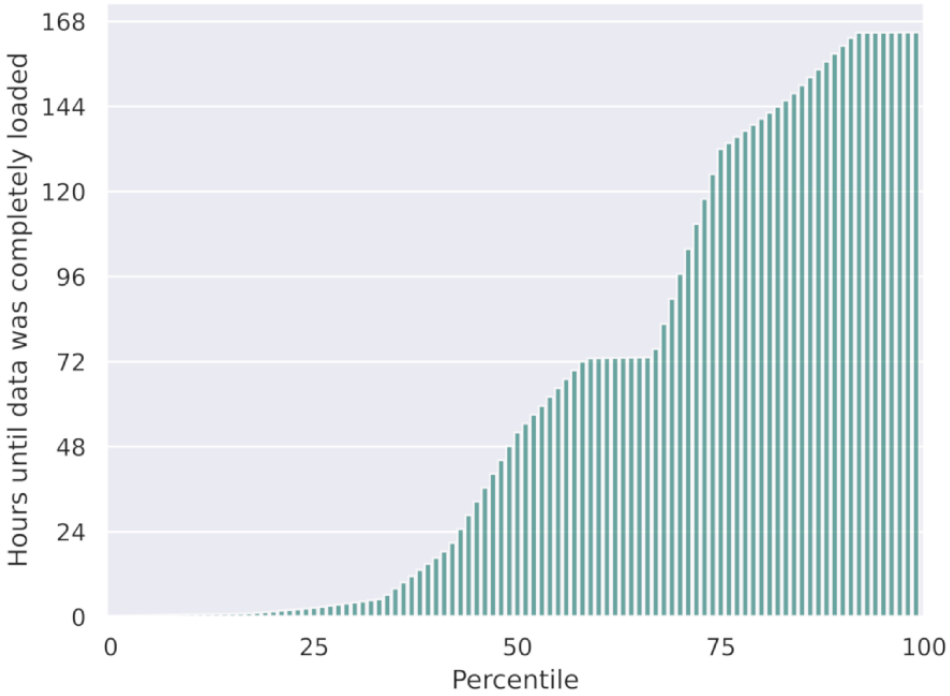


**Figure 3.** The percentiles illustrate all queries conducted by researchers until the availability of data for analysis extracted from the platform logs.




## Discussion

### Principal Findings

The development of a novel data platform at the University Medical Center Hamburg-Eppendorf for hypothesis testing on current clinical routine data according to all European and German data protection laws is accepted and used by clinicians. Accordingly, designing, implementing, and establishing a streamlined process for conducting hypothesis testing in public health by using secondary data appears possible. The initial version presented in this study involves the development of an analytical platform with a data protection-compliant infrastructure and a comprehensive ethical mandate, which will be extended with respect to semantic and syntactic interoperability found in the literature. This innovation has culminated in the establishment of a tool in clinical reality, which occupies a unique niche within the national health care landscape.

Given the illustrative use case, our findings indicate that routine data can facilitate the creation of data sets on scales comparable to prospective studies within significantly shorter time frames than those. This observation carries profound implications for diverse hospital roles: Patients gain transparency and trust in research processes, as the platform serves as a reliable authority for consent, enhancing confidence in the hospital's practices. Clinicians find empirical support for hypothesis testing, aiding in evidence-based decision-making and simplifying time-consuming replication studies. The Data Protection Officer benefits from automated queries, reducing project-related risk management burdens and minimizing infringement risks through a secure, tested architecture. Research data infrastructure experts receive structured support for handling researchers' queries. Finally, the hospital itself benefits from the efficient use of routine clinical data, offering potential cost savings, increased efficiency, and enhanced competitiveness.

Beyond the scope of our study, there is a discernible increase in interest in the tool within clinical reality. Over the last 90 days, the number of successful queries has doubled, and in total, clinical data from 600,000 patients or 1.6 million cases were retrieved from the platform. Although the absolute figures remain constrained, there is evident adoption by clinical researchers, indicating active use of the new tool for their hypothesis tests.

### Limitations

Our investigation underscores that the long execution times of queries on general-purpose databases in clinical systems, which are not inherently designed for the queries executed by the research platform, can limit the interactivity of researchers with the clinical data. Similar systems in other hospitals may likely face comparable issues. The complexity of supporting various query formulations through SQL query adapters further complicates optimization, often resulting in less efficient query statements compared with those that are meticulously crafted by hand. To enhance our platform and achieve shorter execution times, further development is essential. By now, we established a time limitation for queries, terminating excessively large ones. In the future, we plan to use strategies such as horizontally scaling the data sources, using alternative data stores or data caches, or using FHIR search or the Clinical Query Language as the query mechanism instead of traditional SQL [31].

Furthermore, our observations indicate a lack of universal intuitiveness among clinical users in our hospital regarding the Jupyter Notebooks used for analysis. Despite the formulation of data queries, the execution of analyses experienced a notable decline. The participation of clinicians in platform design underscores a potential gap in data literacy among individual

physicians. To mitigate this, we advocate for an additional reduction in the entry barrier through the introduction of user-friendly, broadly applicable dashboards and visualizations tailored to each data query.

An additional aspect that holds potential for enhancing usability in the future is the ability to share access to analysis spaces. This feature would enable users with limited statistical expertise to invite statistical or biomedical experts into their analysis space, gradually receiving support throughout the analysis process. By allowing collaborative access, inexperienced users can benefit from the guidance and assistance of domain experts, facilitating their learning and development in statistical analysis. Accordingly, this feature is currently under development.

## Conclusions

With the presented research platform, we were able to establish a valuable tool for hypothesis testing and secondary use of clinical data. By automating the retrieval process of pseudonymized clinical data and providing a clear legal framework, the platform contributes to the facilitation of the research process. The practical usability of the platform was demonstrated through the replication of a scientific study using the example of PD, confirming the validity of the concept. In further development stages and through the integration of additional clinical data sources, we aim to continuously increase the quantity of data and the usability of the platform. In the long term, through further modularization and standardization, the platform should be made usable for additional national and European sites, significantly facilitating the secondary use of clinical data.

## Acknowledgments

The authors received no direct funding for this work. However, we gratefully acknowledge the financial support provided to the Institute for Applied Medical Informatics by the Dean's Office of the University Medical Center Hamburg Eppendorf, which made this work possible. Furthermore, the work was made feasible through the assistance of the SMITH team at the Hamburg Eppendorf site, which is part of the SMITH consortium funded by the German Federal Ministry of Education and Research (BMBF; grant 01ZZ1803O). We also acknowledge financial support from the Open Access Publication Fund of UKE (Universitätsklinikum Hamburg-Eppendorf) and DFG (German Research Foundation).

During the preparation of this manuscript, the authors used (generative) artificial intelligence–powered tools like DeepL Translator, Grammarly (Grammarly Inc), and ChatGPT (Open AI) to improve the language and style of some parts of the paper. These tools were not used to generate any content of the paper itself.

## Data Availability

The anonymized data sets regarding the usage behavior analyzed during this study are available in the ZFDM repository [32]. The clinical data sets are not publicly available due to the missing legal foundation for the export of routine data but are available from the corresponding author on reasonable request and given permission from the required access committees.

## Authors' Contributions

CG and KG wrote the manuscript. CG and AJW conducted the analysis. MA, MW, and JG led the technical development. FÜ, MW, and JG edited the manuscript. CG is the corresponding author of this paper.

## Conflicts of Interest

None declared.

## References

1. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
2. Semler SC, Wissing F, Heyder R. German medical informatics initiative. Methods Inf Med 2018;57(S 01):e50-e56 [FREE Full text] [doi: 10.3414/ME18-03-0003] [Medline: 30016818]
3. Gehring S, Eulenfeld R. German medical informatics initiative: unlocking data for research and health care. Methods Inf Med 2018;57(S 01):e46-e49 [FREE Full text] [doi: 10.3414/ME18-13-0001] [Medline: 30016817]
4. Medizin Informatik Initiative. Der kerndatensatz der medizininformatik-initiative. URL: https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative [accessed 2023-11-23]
5. Nerius M, Fink A, Doblhammer G. Parkinson's disease in Germany: prevalence and incidence based on health claims data. Acta Neurol Scand 2017;136(5):386-392 [FREE Full text] [doi: 10.1111/ane.12694] [Medline: 27726128]
6. Choe C, Petersen E, Lezius S, Cheng B, Schulz R, Buhmann C, et al. Association of lipid levels with motor and cognitive function and decline in advanced Parkinson's disease in the Mark-PD study. Parkinsonism Relat Disord 2021;85:5-10. [doi: 10.1016/j.parkreldis.2021.02.007] [Medline: 33636481]

7. Soto GE, Spertus JA. EPOCH® and ePRISM®: a web-based translational framework for bridging outcomes research and clinical practice. 2007 Computers in Cardiology 2007:205-208 [FREE Full text] [doi: 10.1109/cic.2007.4745457]
8. Gruendner J, Schwachhofer T, Sippl P, Wolf N, Erpenbeck M, Gulden C, et al. KETOS: clinical decision support and machine learning as a service - a training and deployment platform based on docker, OMOP-CDM, and FHIR web services. PLoS One 2019;14(10):e0223010 [FREE Full text] [doi: 10.1371/journal.pone.0223010] [Medline: 31581246]
9. Valenzuela W, Balsiger F, Wiest R, Scheidegger O. Medical-blocks-a platform for exploration, management, analysis, and sharing of data in biomedical research: system development and integration results. JMIR Form Res 2022;6(4):e32287 [FREE Full text] [doi: 10.2196/32287] [Medline: 35232718]
10. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015;53:162-173 [FREE Full text] [doi: 10.1016/j.jbi.2014.10.006] [Medline: 25463966]
11. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 2013;51(8 Suppl 3):S30-S37 [FREE Full text] [doi: 10.1097/MLR.0b013e31829b1dbd] [Medline: 23774517]
12. Kaggle: Your Machine Learning and Data Science Community. URL: https://www.kaggle.com/ [accessed 2023-06-01]
13. Gradient MLOps Platform. URL: https://www.paperspace.com/gradient [accessed 2023-06-01]
14. Google Colaboratory. URL: https://colab.research.google.com/ [accessed 2023-06-01]
15. CodaLab. URL: https://codalab.org/ [accessed 2023-06-01]
16. Winter A, Stäubert S, Ammon D, Aiche S, Beyan O, Bischoff V, et al. Smart medical information technology for healthcare (SMITH). Methods Inf Med 2018;57(S 01):e92-e105 [FREE Full text] [doi: 10.3414/ME18-02-0004] [Medline: 30016815]
17. Prokosch HU, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical informatics in research and care in university medicine. Methods Inf Med 2018;57(S 01):e82-e91 [FREE Full text] [doi: 10.3414/ME17-02-0025] [Medline: 30016814]
18. § 12 HmbKHG Hamburgisches Krankenhausgesetz (HmbKHG), Gesetze des Bundes und der Länder. 1991. URL: https://www.landesrecht-hamburg.de/bsha/document/jlr-KHGHArahmen [accessed 2023-06-01]
19. Kadioglu D, Breil B, Knell C, Lablans M, Mate S, Schlue D, et al. Samply.MDR – A metadata repository and its application in various research networks. In: German Medical Data Sciences: A Learning Healthcare System. Amsterdam, Netherlands: IOS Press; 2018:50-54.
20. DeMaagd G, Philip A. Parkinson's disease and its management: part 1: disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis. P T 2015;40(8):504-532 [FREE Full text] [Medline: 26236139]
21. Rizek P, Kumar N, Jog MS. An update on the diagnosis and treatment of Parkinson disease. CMAJ 2016;188(16):1157-1165 [FREE Full text] [doi: 10.1503/cmaj.151179] [Medline: 27221269]
22. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, Movement Disorder Society UPDRS Revision Task Force. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
23. Bhatia KP, Bain P, Bajaj N, Elble RJ, Hallett M, Louis ED, Tremor Task Force of the International Parkinson and Movement Disorder Society. Consensus statement on the classification of tremors. from the task force on tremor of the international Parkinson and movement disorder society. Mov Disord 2018;33(1):75-87 [FREE Full text] [doi: 10.1002/mds.27121] [Medline: 29193359]
24. Lenka A, Jankovic J. Tremor syndromes: an updated review. Front Neurol 2021;12:684835 [FREE Full text] [doi: 10.3389/fneur.2021.684835] [Medline: 34381412]
25. Gupta DK, Marano M, Zweber C, Boyd JT, Kuo SH. Prevalence and relationship of rest tremor and action tremor in Parkinson's disease. Tremor Other Hyperkinet Mov (N Y) 2020;10:58 [FREE Full text] [doi: 10.5334/tohm.552] [Medline: 33384882]
26. Louis ED, Levy G, Côte LJ, Mejia H, Fahn S, Marder K. Clinical correlates of action tremor in Parkinson disease. Arch Neurol 2001;58(10):1630-1634. [doi: 10.1001/archneur.58.10.1630] [Medline: 11594921]
27. Parkinson Progression Marker Initiative. The Parkinson progression marker initiative (PPMI). Prog Neurobiol 2011;95(4):629-635 [FREE Full text] [doi: 10.1016/j.pneurobio.2011.09.005] [Medline: 21930184]
28. Kang U, Alcalay R, Goldman J, Henchcliffe C, Hogarth P, Tuite P, et al. The BioFIND study (Fox investigation For new discovery of biomarkers In Parkinson's disease): design and methodology (P4.043).: Neurology Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology; 2014. URL: https://n.neurology.org/content/82/10_Supplement/P4.043 [accessed 2023-06-01]
29. Rosenthal LS, Drake D, Alcalay RN, Babcock D, Bowman FD, Chen-Plotkin A, PDBP consortium. The NINDS Parkinson's disease biomarkers program. Mov Disord 2016;31(6):915-923 [FREE Full text] [doi: 10.1002/mds.26438] [Medline: 26442452]
30. Jagadeesan AJ, Murugesan R, Vimala Devi S, Meera M, Madhumala G, Vishwanathan Padmaja M, et al. Current trends in etiology, prognosis and therapeutic aspects of Parkinson's disease: a review. Acta Biomed 2017;88(3):249-262 [FREE Full text] [doi: 10.23750/abm.v88i3.6063] [Medline: 29083328]

31. Gruendner J, Deppenwiese N, Folz M, Köhler T, Kroll B, Prokosch H, et al. The architecture of a feasibility query portal for distributed COVID-19 fast healthcare interoperability resources (FHIR) patient data repositories: design and implementation study. JMIR Med Inform 2022;10(5):e36709 [FREE Full text] [doi: 10.2196/36709] [Medline: 35486893]
32. Gundler C. Dataset for unlocking the potential of secondary data for public health research: a retrospective study with a novel clinical platform.: ZFDM Repository; 2023. URL: https://www.fdr.uni-hamburg.de/record/13839 [accessed 2024-09-04]

## Abbreviations

**BioFIND:** Fox Investigation for New Discovery of Biomarkers
**CSV:** comma-separated values
**DIC:** data integration center
**EHR4CR:** Electronic Health Records for Clinical Research
**ETL:** Extract, Transform, Load
**FHIR:** Fast Healthcare Interoperability Resources
**HL7:** Health Level Seven International
***ICD-10-GM*:** *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, German Modification*
**MDR:** Metadata Repository
**MDS:** International Parkinson and Movement Disorder Society
**MII:** Medical Informatics Initiative
**PD:** Parkinson disease
**PDBP:** Parkinson's Disease Biomarkers Program
**PPMI:** Parkinson Progression Marker Initiative
**REST:** Representational State Transfer
**SQL:** Structured Query Language
**UKE:** University Medical Center Hamburg-Eppendorf
**UPDRS:** Unified Parkinson's Disease Rating Scale

*Edited by T de Azevedo Cardoso; submitted 03.08.23; peer-reviewed by C Gulden, T Karen, K Fultz Hollis; comments to author 05.10.23; revised version received 01.12.23; accepted 17.07.24; published 01.10.24.*

*Please cite as:*
*Gundler C, Gottfried K, Wiederhold AJ, Ataian M, Wurlitzer M, Gewehr JE, Ückert F*
*Unlocking the Potential of Secondary Data for Public Health Research: Retrospective Study With a Novel Clinical Platform*
*Interact J Med Res 2024;13:e51563*
*URL: https://www.i-jmr.org/2024/1/e51563*
*doi:10.2196/51563*
*PMID:*

Original Paper

# Shift in Demographic Involvement and Clinical Characteristics of COVID-19 From Wild-Type SARS-CoV-2 to the Delta Variant in the Indian Population: In Silico Analysis

Ashutosh Kumar[1], MD; Adil Asghar[1], MD; Khursheed Raza[2], MD; Ravi K Narayan[3], MD; Rakesh K Jha[1], MD; Abhigyan Satyam[1], MD; Gopichand Kumar[1], MBBS; Prakhar Dwivedi[1], MBBS; Chetan Sahni[4], MD; Chiman Kumari[5], MD; Maheswari Kulandhasamy[6], MD; Rohini Motwani[7], MD; Gurjot Kaur[8], PhD; Hare Krishna[9], MD; Sujeet Kumar[10], PhD; Kishore Sesham[11], MD; Sada N Pandey[12], PhD; Rakesh Parashar[13], MD, PhD; Kamla Kant[14], MD

[1]Department of Anatomy, All India Institute of Medical Sciences-Patna, Patna, India
[2]Department of Anatomy, All India Institute of Medical Sciences-Deoghar, Deoghar, Jharkhand, India
[3]Department of Anatomy, All India Institute of Medical Sciences-Bhubaneshwar, Bhubaneshwar, India
[4]Department of Anatomy, All India Institute of Medical Sciences-Gorakhpur, Gorakhpur, India
[5]Department of Anatomy, Postgraduate Institute of Medical Education and Research, Chandigarh, India
[6]Department of Biochemistry, All India Institute of Medical Sciences-Madurai, Madurai, India
[7]Department of Anatomy, All India Institute of Medical Sciences-Bibinagar, Bibinagar, Telangna, India
[8]Department cum National Centre for Human Genome Studies and Research, Punjab University, Chandigarh, India
[9]Department of Anatomy, All India Institute of Medical Sciences-Jodhpur, Jodhpur, Rajasthan, India
[10]School of Allied Health Sciences (Nagpur), Datta Meghe Institute of Higher Education and Research, Wardha, Maharashtra, India
[11]Department of Anatomy, All India Institute of Medical Sciences-Mangalagiri, Mangalagiri, Andhra Pradesh, India
[12]Department of Zoology, Banaras Hindu University, Varanasi, Uttar Pradesh, India
[13]India Health Lead, Oxford Policy Management Limited, Oxford, United Kingdom
[14]Department of Microbiology, All India Institute of Medical Sciences-Bathinda, Bathinda, India

**Corresponding Author:**
Kamla Kant, MD
Department of Microbiology
All India Institute of Medical Sciences-Bathinda
Mandi Dabwali Rd
Bathinda, 151001
India
Phone: 91 0164 286 ext 8710
Email: drkamlakant@gmail.com

## Abstract

**Background:** The Delta variant (B.1.617.2) was considered the most dangerous SARS-CoV-2 strain; however, in-depth studies on its impact based on demographic and clinical characteristics of COVID-19 are scarce.

**Objective:** We aimed to investigate the shift in demographic and clinical characteristics of the COVID-19 pandemic with the emergence of the SARS-CoV-2 Delta variant compared with the wild-type (WT) strain (B.1).

**Methods:** A cross-sectional study of COVID-19 cases in the Indian population caused by the WT strain (B.1) and Delta variant of SARS-CoV-2 was performed. The viral genomic sequence metadata containing demographic, vaccination, and patient status details (N=9500, $N_{Delta}$=6238, $N_{WT}$=3262) were statistically analyzed.

**Results:** With the Delta variant, in comparison with the WT strain, a higher proportion of young individuals (<20 years) were infected (0-9 years: Delta: 281/6238, 4.5% vs B.1: 75/3262, 2.3%; 10-19 years: Delta: 562/6238, 9% vs B.1: 229/3262, 7%; *P*<.001). The proportion of women contracting infection increased (Delta: 2557/6238, 41% vs B.1: 1174/3262, 36%; *P*<.001). However, it decreased for men (Delta: 3681/6238, 59% vs B.1: 2088/3262, 64%; *P*<.001). An increased proportion of the young population developed symptomatic illness and were hospitalized (Delta: 27/262, 10.3% vs B.1: 5/130, 3.8%; *P*=.02). Moreover, an increased proportion of the women (albeit not men) from the young (Delta: 37/262, 14.1% vs B.1: 4/130, 3.1%; *P*<.001) and

adult (Delta: 197/262, 75.2% vs B.1: 72/130, 55.4%; *P*<.001) groups developed symptomatic illness and were hospitalized. The mean age of men and women who contracted infection (Delta: men=37.9, SD 17.2 years; women=36.6, SD 17.6 years; *P*<.001; B.1: men=39.6, SD 16.9 years; women=40.1, SD 17.4 years; *P*<.001) as well as developing symptoms or being hospitalized (Delta: men=39.6, SD 17.4 years; women=35.6, SD 16.9 years, *P*<.001; B.1: men=47, SD 18 years; women=49.5, SD 20.9 years, *P*<.001) were considerably lower with the Delta variant than the B.1 strain. The total mortality was about 1.8 times higher with the Delta variant than with the WT strain. With the Delta variant, compared with B.1, mortality decreased for men (Delta: 58/85, 68% vs B.1: 15/20, 75%; *P*<.001); in contrast, it increased for women (Delta: 27/85, 32% vs B.1: 5/20, 25%; *P*<.001). The odds of death increased with age, irrespective of sex (odds ratio 3.034, 95% CI 1.7-5.2, *P*<.001). Frequent postvaccination infections (24/6238) occurred with the Delta variant following complete doses.

**Conclusions:** The increased involvement of young people and women, the lower mean age for illness, higher mortality, and frequent postvaccination infections were significant epidemiological concerns with the Delta variant.

***(Interact J Med Res 2024;13:e44492)*** doi:10.2196/44492

**KEYWORDS**

SARS-CoV-2; COVID-19; epidemiology; demographic shift; severity of illness; variant; virus; pandemic; population studies; genomic analysis

## Introduction

The Delta variant (B.1.617.2) of SARS-CoV-2 caused COVID-19 waves and spikes in 2021 in multiple countries [1]. Structural and functional analyses of the lineage characterizing mutations in the spike glycoprotein have predicted potential alterations in virus-host interactions and masking of the antibody binding sites, leading to increased transmissibility, lethality, and immune escape capabilities for this variant [2-4], which were further validated in recent animal models [5,6] and human studies [7-10]. Available studies precisely indicate that the Delta variant is at least 50% to 60% more transmissible than the Alpha variant (B.1.1.7) [11] and is capable of immune escape against natural infections with previous SARS-CoV-2 strains, COVID-19 vaccines, and therapeutically used monoclonal antibodies [5-11].

The emerging SARS-CoV-2 variants reportedly vary in demographic characteristics, such as age- and sex-based vulnerability for contracting infection, developing severe illness, and mortality risk from wild-type (WT) strains (Wuhan strain and B.1) [4,12-15]. Against the established pattern of higher susceptibility for older adults and men, which is explained by documented immunological reasons [16-20], the emerging variants involve increasing proportions of the young and female sex [4,12-15]. However, studies that have examined the shift in demographic and clinical characteristics for the Delta variant are currently limited. A devastating second COVID-19 wave occurred in India in 2021, driven by the Delta variant [11,21]. An analysis of the official epidemiological data released by the government of India suggested [21] that a rise in new cases was recorded by February 2021; however, the rate of increase of new cases was slow until the beginning of March and was not reflective of an imminent second wave. A distinct new wave was not in view before the end of March when a sudden spurt was recorded in new cases. The rise in new cases then continued throughout April and peaked by the end of the first week of May 2021, when all global records were surpassed, reporting more than 0.4 million new cases daily. As early as April 2021, nearly 7 million new COVID-19 cases were recorded from the 19 million cases recorded in India during the pandemic. In addition, more than 48,000 deaths occurred in April, of 0.2 million deaths due to COVID-19 reported at that point in India [22]. A sudden rise in cases nationwide [21] shocked the health response system. The first COVID-19 wave, as in other parts of the world, was driven by WT strains in India [21]. The first variant with significantly increased transmissibility and lethality was the Alpha variant, which turned out to be a dominant strain by 2020 [21]. The Alpha variant caused a frequent increase in daily cases across the Indian states. In between, multiple other variants were also reported; however, none of those variants were able to contribute to a significant COVID-19 wave until the arrival of the Delta variant [21]. The first case of the Delta variant was reported in India by December 2020 [21]; however, an exclusive rise in Delta cases was only evident by the end of March 2021 [21]. The Delta variant–driven second wave continued for months, only seeing a decline in July 2021 [21].

As per the official estimates in September 2022, more than 3.4 million cases and 0.52 million deaths were recorded in India, and more than 89 million COVID-19 tests (632,930 per million population) and 200 million vaccinations (at least a single dose) were recorded in India [23]. Notably, the second wave contributed significantly to the total cases and deaths [21].

Plausibly higher transmissibility and virulence of the Delta variant and a lack of substantial vaccination at the time of the second wave could have been the likely reasons for the massive surge in new cases [4,11,21].

This study analyzed the shift in demographic and clinical characteristics of the COVID-19 pandemic in the Indian population with the emergence of the SARS-CoV-2 Delta variant compared with the WT strain (B.1).

## Methods

### Data Collection and Processing

A cross-sectional study was conducted to study the critical differences in demographic and clinical characteristics of the SARS-CoV-2 B.1.617.2 (Delta) variant infections compared with those caused by the WT B.1 strain (Wuhan strain with the D614G mutation) in the Indian population. Genomic sequencing

of the virus strain isolated from patient samples is a standard method for confirmatory diagnosis of COVID-19 and identification of the causative SARS-CoV-2 variant. We analyzed the patient metadata information attached to the genomic sequence reports for the SARS-CoV-2 Delta and WT (B.1) strains detected in the patient samples from the Indian population. The SARS-CoV-2 genomic sequence reports from India (patient sample collection date no later than July 31, 2021) were accessed from the EpiCoV database of the Global Initiative on Sharing All Influenza Data (GISAID) [24] using the automatic search function for information by geographical location, SARS-CoV-2 lineage, and collection dates. A similar search strategy was used to retrieve comparative metadata for the COVID-19 cases with the WT B.1 strain reported on GISAID since the first COVID-19 wave. The metadata files of the Delta variant (test group) and WT (control group) strain sequences were downloaded from the individual GISAID accession numbers, checked and confirmed for accuracy of the lineage information, and screened for demographic details (age and sex) and vaccination status (for 2 complete doses). Further, the clinical outcomes (information about the illness severity and mortality during the disease course) of the WT and Delta variant strains cases were assessed using a repeat search entering "patient status" as additional input. The collected data were filtered by discarding the sequence reports containing no or incomplete information and duplications (repeat sequencing from the same individuals). To prevent selection bias in the data collection, no specific filters for age, sex, date of collection, and geographical location were used during the search. A team of 3 investigators rechecked and verified the sampling and data entry errors of the complete data set by matching the original details in the metadata files.

The sample size for the study was guided by serosurvey reports from the Indian population during successive waves of the pandemic [25,26]. Of note, nationwide seroprevalences of 24.1% and 67.6% were reported in the studies conducted at the end of the first and second waves, respectively [25,26]. Considering a minimum of 24.1% seroprevalence in the current population of India (approximately 1.35 billion) at 5% relative precision (1.25% absolute precision) and 95% CI, a minimum sample size of 4866 was determined for the study.

### Data Analysis and Presentation

#### *Quantitative Variables*

The final data were entered in Excel sheets, and the distribution was analyzed by age and sex. The age distribution of the Delta variant infections was presented at 10-year intervals. Additionally, for the analysis of clinical outcomes, all cases were divided into 3 broad age groups as follows: "young" (0-19 years), "adult" (20-59 years), and "older adult" (≥60 years).

#### *Clinical Outcomes*

Clinical outcomes of the positive cases were categorized as "asymptomatic/mildly symptomatic" (including cases in home isolation and/or quarantine with no overt symptoms or mild symptoms), "symptomatic/hospitalized" (including cases with overt symptoms, currently hospitalized, or released or recovered after hospitalization), and "demised." These were also the significant outcomes assessed in this study.

#### *Data Distribution and Analysis, Statistical Assessment, and Graphical Presentation*

Individual outcomes were analyzed by age and sex distribution. To determine mortality, the statuses of all the categories other than "demised" were considered "living." Categorical data are presented as frequency or proportion. Continuous data are presented as mean (or median) and SD. The relative contribution of the geographical locations within India (states and union territories) in the total available genomic data from India in the study period was noted; however, no further distinction of specific geographical locations was made while analyzing the final data.

Statistical tests were performed to evaluate intergroup differences in clinical outcomes with the help of Microsoft Excel 2019 and the XLSTAT package. The normality of the data was examined using the Kolmogorov-Smirnov test. For normally distributed data, 2-sample Student *t* tests and ANOVAs were used. The Games-Howell post hoc test was applied for intergroup comparisons. The Kruskal Wallis H test was used to analyze skewed data. The chi-square test was used for categorical variables. Multinomial logistic regression analysis was used to estimate the relative measures of effect, considering "demised" and "symptomatic/hospitalized" as the clinical outcomes in reference to the Delta variant versus the WT (B.1) strain. No specific theoretical nor statistical criteria were adopted for "age"; however, "B.1" for the virus strain and "men" for sex were set as the baseline or reference in the regression model.

Results were considered statistically significant at a $P$ value ≤.05. Graphs were plotted to visualize the data trends.

### Ethical Considerations

Approval from the institute ethics committee was precluded, as the study used open-access data from the EpiCoVTM database from the GISAID [24]. GISAID permits open access to data to all individuals who agree to identify themselves and follow the GISAID sharing mechanism governed by its Database Access Agreement [27].

## Results

### Incidence

We assessed the genomic sequence reports for 8269 and 3767 cases with the Delta variant and WT B.1 strains, respectively, that were uploaded on GISAID from the first case to July 31, 2021. The SARS-CoV-2 genomic sequences were noted from nearly all the Indian states and union territories (Figure S1 in Multimedia Appendix 1). After filtering the data, 6238 Delta variant and 3262 WT B.1 strain cases were available for demographic analysis.

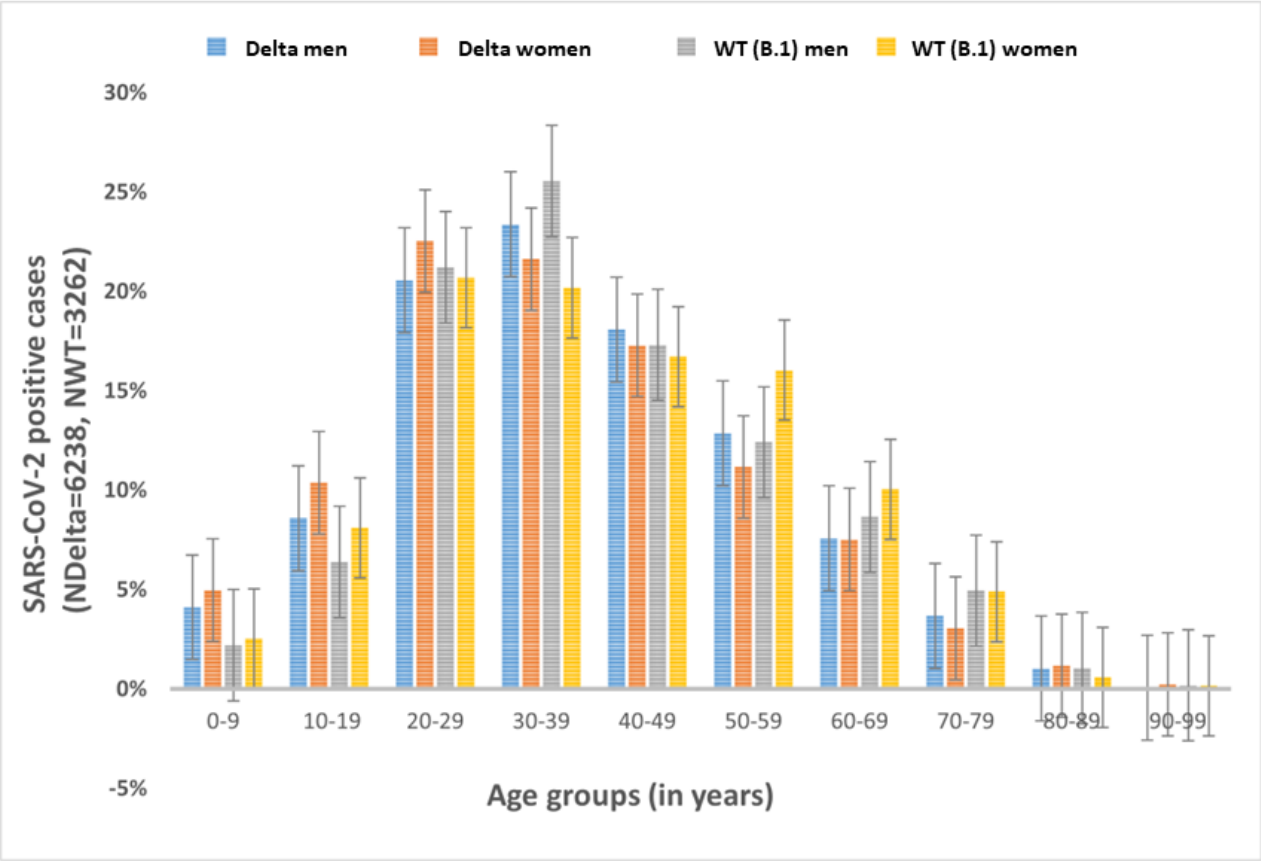
Among the screened cases, there were 24 Delta infections following 2 complete vaccine doses; however, no postvaccination infections were noted with B.1. In the database, 659 and 320 sequence reports with information on "patient status" were present for Delta and B.1 strains, respectively, and after filtering the data, 647 and 276 cases, respectively, were

available for analysis regarding illness severity and mortality in terms of age and sex.

## Demographic Distribution

Frequent incidences of Delta variant as well as B.1 strain infections (N=6238) were noted across the age groups, including in the young group (0-19 years; Figure 1). The prevalence of infections was higher with the Delta variant than with the B.1 strain in the young group (0-19 years), with the highest difference in prevalence in those aged 0 years to 9 years followed by those aged 10 years to 19 years (*P*<.001). Conversely, the differences in prevalences were relatively low in those aged 20 years to 60 years, with a few exceptions (*P*<.001; Figure 1). Although a greater proportion of men were infected with both variants (both sexes: $N_{Delta}$=6238; $N_{WT}$ =3262) than women (Delta: 3681/6238, 59% vs 2557/6238, 41%; WT: 2088/3262, 64% vs 1174/3262, 36%), the proportion of women contracting infection was higher with Delta than with B.1 (Delta: 2557/6238, 41% vs B.1: 1174/3262, 36%; *P*<.001). The mean ages when infected with the Delta variant were 37.9 (SD 17.2) years for men and 36.6 (SD 17.6) years for women and when infected with the B.1 strain were 39.6 (SD 16.9) years for men and 40.1 (SD 17.4) years for women (*P*<.001).

**Figure 1.** Demographic distribution of SARS-CoV-2 B.1.617.2 (Delta) variant versus the wild type (WT) strain (B.1) infections in Indian population.




## Patient Status

The cases were noted for all categories for the Delta variant as well as B.1 (asymptomatic/mildly symptomatic: $N_{Delta}$=72, $N_{WT}$=24; symptomatic/hospitalized: $N_{Delta}$=262, $N_{WT}$=130; living but symptom status not known: $N_{Delta}$=228, $N_{WT}$=102; demised: $N_{Delta}$ =85, $N_{WT}$=20). The demographic (age groups and sex) distributions by patient status are shown in Figure S2 in Multimedia Appendix 1.

### *Symptomatic Illness and Hospitalization*

The cases who developed symptoms and required hospitalization were reported in each age group, including the young group (0-19 years) for both strains (Figure 2A). However, the young group (0-19 years; Delta: 27/262, 10.3% vs B.1: 5/130, 3.8%; *P*=.02) contributed to the increased proportion of the total number of symptomatic or hospitalized cases with the Delta variant. Further, higher proportions of young (Delta: 37/262, 14.1% vs B.1: 4/130, 3.1%) and adult (Delta: 197/262, 75.2% vs B.1: 72/130, 55.4%) women, but not men, developed symptoms and required hospitalization with the Delta variant than with B.1 (*P*<.001; Figure 2B). The mean ages at developing symptoms or hospitalization for men and women were 39.6 (SD 17.4) years and 35.6 (SD 16.9) years, respectively, with the Delta variant and 47 (SD 18) years and 49.5 (SD 20.9) years, respectively, with the B.1 strain (*P*<.001). However, in the multinomial logistic regression model, the risk of symptomatic illness or hospitalization was marginally higher for those of a lower age, with borderline statistical significance (odds ratio [OR] –1.1992, 95% CI 0.95-0.99; *P*=.046) for both strains. The viral lineage (OR 0.695, 95% CI 0.415-1.166; *P*=.17) and sex (OR 1.052, 95% CI 0.659-1.679; *P*=.83) had no significant effect on the estimates (Table 1).

**Figure 2.** Distribution of symptomatic or hospitalized cases with SARS-CoV-2 B.1.617.2 (Delta) variant or wild-type (WT) strain (B.1) infections by (A) age (*P*=.02; statistical significance set at *P*<.05) and (B) sex (*P*<.001; statistical significance set at *P*<.001).

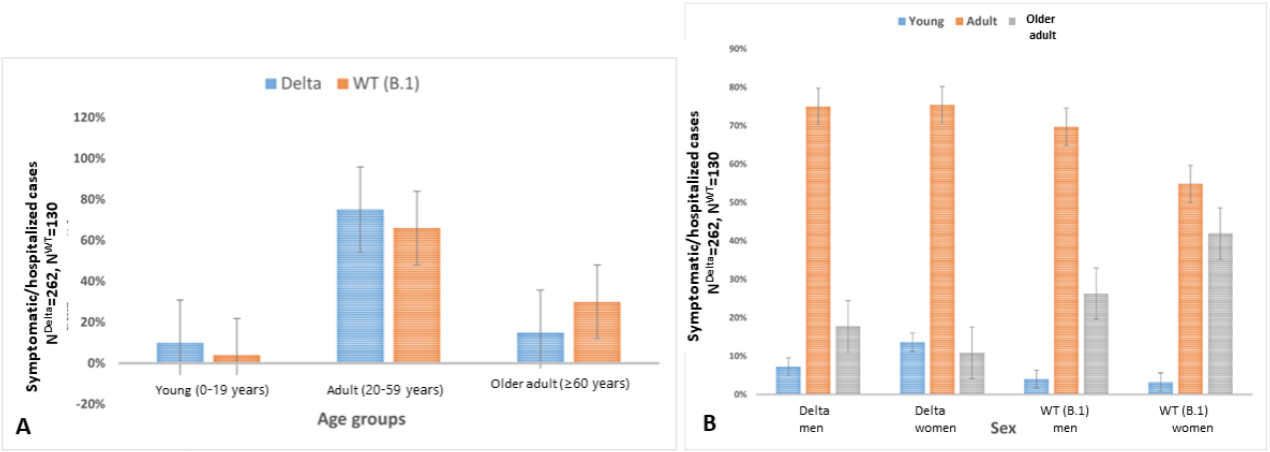



**Table 1.** The viral and host factors influencing the clinical outcomes of death and symptomatic illness or hospitalization.

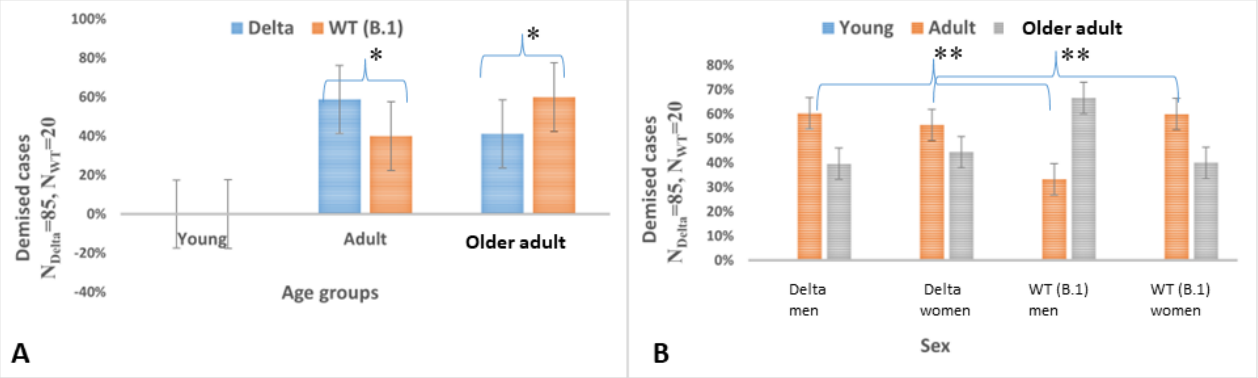
| Clinical outcome and predictors | *z* distribution | *P* value | Odds ratio (95% CI) |
|---|---|---|---|
| **Death** | | | |
| SARS-CoV-2 lineage: B.1[a]-Delta | 3.93 | <.001 | 3.034 (1.744-5.277) |
| Age | –9.10 | <.001 | 0.934 (0.920-0.947) |
| Sex: men[a]-women | –1.11 | .27 | 0.765 (0.477-1.227) |
| **Symptomatic illness or hospitalized** | | | |
| SARS-CoV-2 lineage: B.1[a]-Delta | –1.378 | .17 | 0.695 (0.415-1.166) |
| Age | –1.992 | .046 | 0.977 (0.956-0.999) |
| Sex: men[a]-women | 0.213 | .83 | 1.052 (0.659-1.679) |

[a]Reference predictors: SARS-CoV-2 lineage, age, and sex.

### *Total Mortality and Odds of Death*

Of the total cases, approximately 13.1% (85/647) and 7.2% (20/276) died due to COVID-19 caused by the Delta and B.1 strains, respectively. However, no deaths were reported for the Delta variant nor B.1 strain in those aged <20 years (Figure 3A). Further, the adult group contributed to a higher mortality with Delta variant infections than with B.1 strain infections (50/85, 59% vs 8/20, 40%; *P*=.01). When the adult and older adult age groups were analyzed together, there was greater mortality for men than women with the Delta variant (men: 58/85, 68% vs women: 27/85, 32%) and B.1 strain (men: 15/20, 75% vs women: 5/20, 25%; *P*<.001; Figure 3B). Notably, mortality with the Delta variant was lower for men (Delta: 58/85, 68% vs B.1: 15/20, 75%) but higher for women (Delta: 27/85, 32% vs B.1: 5/20, 25%) than with the B.1 strain (*P*<.001). The mean ages at illness-related mortality for men and women were 56.6 (SD 13.5) years and 58.8 (SD 11.4) years, respectively, with the Delta variant and 60.7 (SD 15.5) years and 50.2 (SD 15.7) years, respectively, with the B.1 strain (*P*<.001). The odds of death were higher with Delta variant infections irrespective of sex (OR 3.034, 95% CI 1.7-5.2; *P*<.001). Age but not sex influenced the effect estimates. Lower age was marginally protective (OR 0.934, 95% CI 0.92-0.95; *P*<.001) with both strains (Table 1).

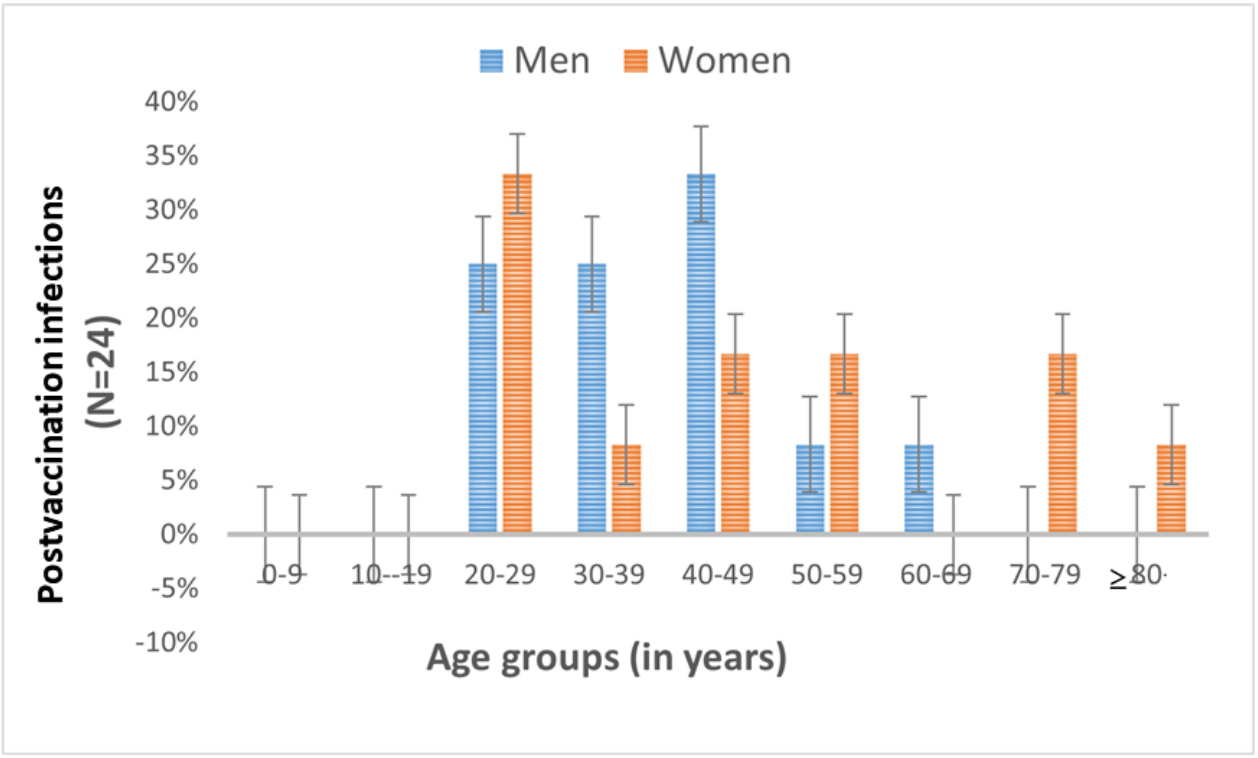
**Figure 3.** Distribution of demised cases with SARS-CoV-2 B.1.617.2 (Delta) variant versus wild type (WT) strain (B.1) infections by (A) age and (B) sex. *Statistical significance set at *P*<.05; **statistical significance set at *P*<.001.




## Postvaccination Infections

Of all the screened cases with the Delta variant or B.1 strain, Delta infections following 2 complete doses of vaccines did occur (24/6238: men=12, women=12); however, no postvaccination infections were noted with B.1 (0/3262). Postvaccination infections with the Delta variant were reported across the age groups (except for those aged 0-9 years or 10-19 years, which is explained by the fact that COVID-19 vaccines were not administered to individuals younger than 18 years in India until the period of this study) and for both sexes (Figure 4). The mean ages at postvaccination infection were 44.48 (SD 16.17) years for all cases, 40.4 (SD 11.9) years for men, and 47.8 (SD 20.4) years for women. No significant differences were noted in the frequency of postvaccination infections in terms of age groups and sex. Of all postvaccination infections, 88% (21/24) were with Covishield [28], and the rest (3/24, 12%) were with Covaxin [29].

**Figure 4.** Demographic distribution of postvaccination (2 complete doses) SARS-CoV-2 B.1.617.2 (Delta) variant infections.




# Discussion

## Principal Findings

Our analyses revealed essential observations indicating a shift in clinical and epidemiological characteristics of the COVID-19 pandemic from the WT strain B.1 to the Delta variant. With the Delta variant, more young individuals (<20 years) and adult women contracted the infection. In addition, an increased proportion of the young (<20 years) and adult population developed symptoms and required hospitalization. The mean age at COVID-19 infection and developing symptomatic illness or hospitalization was considerably lower. A disproportionate number of cases who developed symptomatic illness or were hospitalized was noted by sex: An increased proportion of young and adult women was affected. The mortality rate was about

1.8 times higher. Sex-specific contributions to the mortality data were lower for men but higher for women. However, the odds of death increased with age, irrespective of sex. Further, we noted multiple COVID-19 infections with the Delta variant following 2 complete vaccine doses.

Studies analyzing the demographic characteristics of the Delta variant in the Indian population are currently scarce. Recently, Kumar et al [30] presented their observations on the Indian population using data from the national clinical registry for COVID-19. The authors studied the clinical profiles of hospitalized patients during the second COVID-19 wave, mainly driven by the Delta variant [11,21]. Authors have also studied hospitalized cases in the first COVID-19 wave, primarily caused by WT strains [11,21]. However, the diagnosis of COVID-19 in cases included in this study was based on polymerase chain reaction test results, and no viral genomic sequencing was performed to ensure that the analyzed cases only represented those caused by the Delta variant or WT strains in the respective waves. The authors noted a decrease in the mean age of the total population who contracted infection and a reduction in the proportion of cases who were men during the second wave. The authors noted that mortality among hospitalized patients was 13.26% in the second wave, which was 3.1% higher than in the first wave. The increase in mortality was seen in all age groups in their study except for those aged <20 years, for whom mortality decreased [30]. We observed a similar trend in the susceptibility for contracting infection and risk of mortality in terms of age and sex, except we noted greater total mortality with the Delta variant (13.1%) than with the WT strain (7.2%). However, no deaths with either strain were noted in those aged <20 years in this study.

India is one of the world's most populated countries, with a very high proportion of young people (approximately 45% are <20 years old) [31]. Our analyses show that, in comparison with B.1, Delta variant infections have occurred at considerably higher proportions in those <20 years old, particularly for those aged 0 years to 9 years (Delta: 4.5% vs B.1: 2.3%). The proportion of symptomatic cases and those requiring hospital admission was also higher in the young group (Delta: 10% vs B.1: 4%; Figure 2A). These findings indicate an age shift for the Delta variant, as COVID-19 infections with the WT strains were not common in young people and development of severe symptoms and the need for hospitalization were primarily limited to adults, more specifically older adults [17,18].

We observed that a higher proportion of women than men contracted the infection and developed symptoms or were hospitalized; in addition, mortality was lower in men but higher in women with the Delta variant than with the B.1 strain. Biologically, the severity of illness and mortality risks are considered lower for women than men for infectious diseases that affect mammals, including humans [32]. A higher vulnerability for contracting infection and symptomatic disease for women and narrowing the sex gap in mortality risk with the Delta variant present a paramount epidemiological concern.

Increased involvement of the younger age groups, primarily those aged 0 years to 19 years and of the female sex, in Delta variant infections compared with WT strains was indicated by studies and surveillance reports in the United Kingdom following a surge in Delta variant infections [8,33]. Despite the established immunological advantage for younger individuals [19,20,34] and women [16], as were reflected in infections with WT strains [16,18], increased risks of contracting infection and greater illness severity in these demographic groups indicate enhanced virulence and lethality with the Delta variant.

Last, the multiple Delta variant infections following 2 complete doses of COVID-19 vaccines in our study are supported by previous reports of vaccine breakthrough infections with this variant, including those from the Indian population [9,35].

The findings of this study and others describing demographic and clinical involvement in Delta variant infections confirm a shift from the established pattern for WT strains. Multiple factors could contribute to this epidemiological shift, most importantly, the variant's intrinsic properties, such as increased transmissibility, virulence, and immune escape capabilities, as have been indicated by recent studies [7-9,11,30,36]. A sudden spurt in the number of cases during the second wave and consequent burden on the emergency health response system could have contributed to the increased mortality noted with the Delta variant. In contrast, the prioritized vaccination of those aged >45 years might have protected the vaccinated older adult population and contributed to lowering the mean age for COVID-19 involvement. On the other hand, in the initial period of the first COVID-19 wave, there was a shortage of diagnostic and COVID-19 care facilities and skilled health personnel in the country as well as limited knowledge about the prevention and therapeutic management of the pandemic. When combined, these limitations might have had an impact on the quality of the available epidemiological and clinical data related to WT strains.

## Limitations of the Study

Our study has multiple limitations that must be considered when interpreting the findings. First, our data for reporting postvaccination infections and patient status are limited; hence, related observations may require further validation from studies with a larger sample size. Second, the vaccination status data were unavailable in most of the accessed genomic sequence reports; hence, our data for postvaccination infections do not reflect the incidences of such cases in the total population. Our study does not rule out that actual incidences of postvaccination infections (after 2 complete doses) with the Delta variant could be higher. In addition, the noted 0 postvaccination infections with B.1 in our study could be an underestimation, as COVID-19 vaccines were yet to be made available to the masses in India during the first wave when the WT strains caused most of the infections.

Moreover, owing to the limited availability of vaccination status details for the screened cases in the studied database, we cannot comment conclusively on the efficacy of the vaccines noted in this study. Third, we have not considered the influence of confounding factors, such as comorbidities, vaccination status, and previous COVID-19 infection, on the clinical outcomes of the studied cases, which may have influenced the quality of the results. Last, although we applied no age, sex, nor region-specific filters when acquiring the data, a complete absence of selection bias cannot be assured when uploading the

primary data on GISAID. However, any inherent bias in the primary data will likely be nondifferential and may not significantly impact the comparative analysis between the viral strains.

## Conclusions

In conclusion, our study reveals greater proportions of infected young individuals and women, a lower mean age for illness, greater mortality, and frequent postvaccination infections with the Delta variant than with the WT (B.1) strain. Collectively, the findings of this study elaborate on the changing demographic characteristics of the COVID-19 pandemic with the emergence of the Delta variant. These findings are essential because we have only considered in the analyses those COVID-19 cases for whom Delta variant infections were confirmed by genomic sequencing, ensuring the accuracy of the presented data.

## Acknowledgments

This study used SARS-CoV-2 genomic sequence data from the EpiCoV database of the Global Initiative on Sharing All Influenza Data (GISAID) [24].

## Data Availability

The primary data for this study are publicly available on the SARS-CoV-2 genomic sequence-GISAID database [24]. The tabulated data can be obtained from the corresponding author upon reasonable request.

## Authors' Contributions

AK designed the study and wrote the first draft of the manuscript. KR, RKN, RKJ, AS, GK, and PD performed the data collection and analysis. AA performed the statistical analyses. AA, RKN, CS, CK, MK, RM, GK, HK, KS, SNP, RP, KK, and SK reviewed and edited the manuscript. All authors consented to the submission of the final draft.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of cases geographically and by patient status.
[DOCX File , 727 KB - ijmr_v13i1e44492_app1.docx ]

## References

1. B.1.617.2) Lineage Report. Outbreak Info. URL: https://outbreak.info/situation-reports?pango=B.1.617.2) [accessed 2024-08-31]
2. Khateeb J, Li Y, Zhang H. Emerging SARS-CoV-2 variants of concern and potential intervention approaches. Crit Care 2021 Jul 12;25(1):244 [FREE Full text] [doi: 10.1186/s13054-021-03662-x] [Medline: 34253247]
3. McCallum M, Walls AC, Sprouse KR, Bowen JE, Rosen LE, Dang HV, et al. Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. Science 2021 Dec 24;374(6575):1621-1626 [FREE Full text] [doi: 10.1126/science.abl8506] [Medline: 34751595]
4. Kumar A, Parashar R, Faiq MA, Kumar S, Kumari C, Kulandhasamy M, et al. Emerging SARS-CoV-2 variants can potentially break set epidemiological barriers in COVID-19. SSRN Journal 2021:1. [doi: 10.2139/ssrn.3888058]
5. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature 2021 Aug;596(7871):276-280. [doi: 10.1038/s41586-021-03777-9] [Medline: 34237773]
6. Mohandas S, Yadav PD, Shete A, Nyayanit D, Sapkal G, Lole K, et al. SARS-CoV-2 Delta variant pathogenesis and host response in Syrian hamsters. Viruses 2021 Sep 05;13(9):1773. [doi: 10.3390/V13091773]
7. Li B, Deng A, Li K, Hu Y, Li Z, Shi Y, et al. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. Nat Commun 2022 Jan 24;13(1):460 [FREE Full text] [doi: 10.1038/s41467-022-28089-y] [Medline: 35075154]
8. Twohig KA, Nyberg T, Zaidi A, Thelwall S, Sinnathamby MA, Aliabadi S, COVID-19 Genomics UK (COG-UK) consortium. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. Lancet Infect Dis 2022 Jan;22(1):35-42 [FREE Full text] [doi: 10.1016/S1473-3099(21)00475-8] [Medline: 34461056]
9. Gupta N, Kaur H, Yadav P, Mukhopadhyay L, Sahay RR, Kumar A, et al. Clinical characterization and genomic analysis of samples from COVID-19 breakthrough infections during the second wave among the various states of India. Viruses 2021 Sep 07;13(9):1 [FREE Full text] [doi: 10.3390/v13091782] [Medline: 34578363]

10. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro Surveill 2021 Jun;26(24):1 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.24.2100509] [Medline: 34142653]
11. Dhar MS, Marwal R, Vs R, Ponnusamy K, Jolly B, Bhoyar RC, et al. Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. Science 2021 Nov 19;374(6570):995-999 [FREE Full text] [doi: 10.1126/science.abj9932] [Medline: 34648303]
12. Yang W, Shaman J. Development of a model-inference system for estimating epidemiological characteristics of SARS-CoV-2 variants of concern. Nat Commun 2021 Sep 22;12(1):5573. [doi: 10.1038/s41467-021-25913-9] [Medline: 34552095]
13. Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. Lancet Infect Dis 2021 Oct;21(9):1246-1256 [FREE Full text] [doi: 10.1016/S1473-3099(21)00170-5] [Medline: 33857406]
14. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science 2021 May 21;372(6544):815-821 [FREE Full text] [doi: 10.1126/science.abh2644] [Medline: 33853970]
15. Eurosurveillance editorial team. Updated rapid risk assessment from ECDC on the risk related to the spread of new SARS-CoV-2 variants of concern in the EU/EEA - first update. Euro Surveill 2021 Jan;26(3):2100348 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.3.2101211] [Medline: 33478621]
16. Kumar A, Narayan RK, Kulandhasamy M, Prasoon P, Kumari C, Kumar S, et al. COVID-19 pandemic: insights into molecular mechanisms leading to sex-based differences in patient outcomes. Expert Rev Mol Med 2021 Aug 03;23:e7 [FREE Full text] [doi: 10.1017/erm.2021.9] [Medline: 34340720]
17. Chen Y, Klein SL, Garibaldi BT, Li H, Wu C, Osevala NM, et al. Aging in COVID-19: Vulnerability, immunity and intervention. Ageing Res Rev 2021 Jan;65:101205 [FREE Full text] [doi: 10.1016/j.arr.2020.101205] [Medline: 33137510]
18. Irfan O, Li J, Tang K, Wang Z, Bhutta ZA. Risk of infection and transmission of SARS-CoV-2 among children and adolescents in households, communities and educational settings: A systematic review and meta-analysis. J Glob Health 2021 Jul 17;11:05013 [FREE Full text] [doi: 10.7189/jogh.11.05013] [Medline: 34326997]
19. Pierce C, Sy S, Galen B, Goldstein DY, Orner E, Keller MJ, et al. Natural mucosal barriers and COVID-19 in children. JCI Insight 2021 May 10;6(9):1 [FREE Full text] [doi: 10.1172/jci.insight.148694] [Medline: 33822777]
20. Cohen CA, Li APY, Hachim A, Hui DSC, Kwan MYW, Tsang OTY, et al. SARS-CoV-2 specific T cell responses are lower in children and increase with age and time after infection. Nat Commun 2021 Jul 29;12(1):4678 [FREE Full text] [doi: 10.1038/s41467-021-24938-4] [Medline: 34326343]
21. Kumar A, Asghar A, Dwivedi P, Kumar G, Narayan RK, Jha RK, et al. A bioinformatics tool for predicting future COVID-19 waves based on a retrospective analysis of the second wave in India: model development study. JMIR Bioinform Biotechnol 2022 Jul;3(1):e36860-e36869 [FREE Full text] [doi: 10.2196/36860] [Medline: 36193192]
22. India. John Hopkins University & Medicine Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/region/india [accessed 2024-08-31]
23. Ministry of Health & Family Welfare. URL: https://www.mohfw.gov.in/ [accessed 2024-08-31]
24. GISAID. URL: https://gisaid.org/ [accessed 2024-08-31]
25. Murhekar MV, Bhatnagar T, Thangaraj JWV, Saravanakumar V, Kumar MS, Selvaraju S, et al. SARS-CoV-2 seroprevalence among the general population and healthcare workers in India, December 2020-January 2021. Int J Infect Dis 2021 Jul;108:145-155 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.040] [Medline: 34022338]
26. Dasgupta R. After India's brutal coronavirus wave, two-thirds of population has been exposed to SARS-CoV2. The Conversation. 2021 Aug 20. URL: https://theconversation.com/after-indias-brutal-coronavirus-wave-two-thirds-of-population-has-been-exposed-to-sars-cov2-165050 [accessed 2024-08-31]
27. GISAID EpiFlu™ Database Access Agreement. GISAID. URL: https://gisaid.org/terms-of-use/ [accessed 2024-08-31]
28. Coronavirus disease (COVID-19): Vaccines. Serum Institute of India Pvt Ltd. URL: https://www.seruminstitute.com/product_covishield.php [accessed 2024-08-31]
29. Covaxin. Bharat Biotech. URL: https://www.bharatbiotech.com/covaxin.html [accessed 2024-08-31]
30. Kumar G, Mukherjee A, Sharma R, Menon G, Sahu D, Wig N, National Clinical Registry for COVID-19 Team. Clinical profile of hospitalized COVID-19 patients in first and second wave of the pandemic: Insights from an Indian registry based observational study. Indian J Med Res 2021 May;153(5&6):619-628 [FREE Full text] [doi: 10.4103/ijmr.ijmr_1628_21] [Medline: 34259194]
31. C-02: Marital status by age and sex (total), India - 2011. Census of India. 2021 Jan 19. URL: https://censusindia.gov.in/nada/index.php/catalog/10301 [accessed 2024-08-31]
32. Scully EP, Haverfield J, Ursin RL, Tannenbaum C, Klein SL. Considering how biological sex impacts immune responses and COVID-19 outcomes. Nat Rev Immunol 2020 Jul 11;20(7):442-447 [FREE Full text] [doi: 10.1038/s41577-020-0348-8] [Medline: 32528136]
33. Investigation of SARS-CoV-2 variants of concern: technical briefings. Public Health England. 2021 Sep 17. URL: https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201 [accessed 2024-08-31]

34. Selva KJ, van de Sandt CE, Lemke MM, Lee CY, Shoffner SK, Chua BY, et al. Systems serology detects functionally distinct coronavirus antibody features in children and elderly. Nat Commun 2021 Apr 01;12(1):2037. [doi: 10.1038/s41467-021-22236-7] [Medline: 33795692]
35. Hacisuleyman E, Hale C, Saito Y, Blachere NE, Bergh M, Conlon EG, et al. Vaccine breakthrough infections with SARS-CoV-2 variants. N Engl J Med 2021 Jun 10;384(23):2212-2218. [doi: 10.1056/nejmoa2105000]
36. Kislaya I, Rodrigues E, Borges V, Gomes JP, Sousa C, Almeida JP, PT-Covid-19 Group. Delta variant and mRNA Covid-19 vaccines effectiveness: higher odds of vaccine infection breakthroughs. MedRxiv Preprint posted online on August 22, 2021. [doi: 10.1101/2021.08.14.21262020]

## Abbreviations

**GISAID:** Global Initiative on Sharing All Influenza Data
**OR:** odds ratio
**WT:** wild-type

*Edited by T de Azevedo Cardoso; submitted 21.11.22; peer-reviewed by D Bondage, N Bankar; comments to author 22.08.23; revised version received 04.09.23; accepted 21.06.24; published 08.10.24.*

*Please cite as:*
*Kumar A, Asghar A, Raza K, Narayan RK, Jha RK, Satyam A, Kumar G, Dwivedi P, Sahni C, Kumari C, Kulandhasamy M, Motwani R, Kaur G, Krishna H, Kumar S, Sesham K, Pandey SN, Parashar R, Kant K*
*Shift in Demographic Involvement and Clinical Characteristics of COVID-19 From Wild-Type SARS-CoV-2 to the Delta Variant in the Indian Population: In Silico Analysis*
*Interact J Med Res 2024;13:e44492*
*URL: https://www.i-jmr.org/2024/1/e44492*
*doi:10.2196/44492*
*PMID:*

Original Paper

# Changes in the Epidemiological Features of Influenza After the COVID-19 Pandemic in China, the United States, and Australia: Updated Surveillance Data for Influenza Activity

Mingyue Jiang[1], PhD; Mengmeng Jia[2], PhD; Qing Wang[1], PhD; Yanxia Sun[1], PhD; Yunshao Xu[1], PhD; Peixi Dai[3], MPH; Weizhong Yang[1], PhD; Luzhao Feng[1], PhD

[1]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]National Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[3]Chinese Center for Disease Control and Prevention, Beijing, China

**Corresponding Author:**
Luzhao Feng, PhD
School of Population Medicine and Public Health
Chinese Academy of Medical Sciences & Peking Union Medical College
31 Beiji Tower
Dongcheng District
Beijing, 100730
China
Phone: 86 10 65120716
Email: fengluzhao@cams.cn

## Abstract

**Background:** There has been a global decrease in seasonal influenza activity since the onset of the COVID-19 pandemic.

**Objective:** We aimed to describe influenza activity during the 2021/2022 season and compare it to the trends from 2012 to 2023. We also explored the influence of social and public health prevention measures during the COVID-19 pandemic on influenza activity.

**Methods:** We obtained influenza data from January 1, 2012, to February 5, 2023, from publicly available platforms for China, the United States, and Australia. Mitigation measures were evaluated per the stringency index, a composite index with 9 measures. A general additive model was used to assess the stringency index and the influenza positivity rate correlation, and the deviance explained was calculated.

**Results:** We used over 200,000 influenza surveillance data. Influenza activity remained low in the United States and Australia during the 2021/2022 season. However, it increased in the United States with a positive rate of 26.2% in the 49th week of 2022. During the 2021/2022 season, influenza activity significantly increased compared with the previous year in southern and northern China, with peak positivity rates of 28.1% and 35.1% in the second week of 2022, respectively. After the COVID-19 pandemic, the dominant influenza virus genotype in China was type B/Victoria, during the 2021/2022 season, and accounted for >98% (24,541/24,908 in the South and 20,543/20,634 in the North) of all cases. Influenza virus type B/Yamagata was not detected in all these areas after the COVID-19 pandemic. Several measures individually significantly influence local influenza activity, except for influenza type B in Australia. When combined with all the measures, the deviance explained values for influenza A and B were 87.4% ($P$<.05 for measures of close public transport and restrictions on international travel) and 77.6% in southern China and 83.4% ($P$<.05 for measures of school closing and close public transport) and 81.4% in northern China, respectively. In the United States, the association was relatively stronger, with deviance-explained values of 98.6% for influenza A and 99.1% ($P$<.05 for measures of restrictions on international travel and public information campaign) for influenza B. There were no discernible effects on influenza B activity in Australia between 2020 and 2022 due to the incredibly low positive rate of influenza B. Additionally, the deviance explained values were 95.8% ($P$<.05 for measures of restrictions on gathering size and restrictions on international travel) for influenza A and 72.7% for influenza B.

**Conclusions:** Influenza activity has increased gradually since 2021. Mitigation measures for COVID-19 showed correlations with influenza activity, mainly driven by the early stage of the pandemic. During late 2021 and 2022, the influence of mitigation management for COVID-19 seemingly decreased gradually, as the activity of influenza increased compared to the 2020/2021 season.

*(Interact J Med Res 2024;13:e47370)* doi:10.2196/47370

**KEYWORDS**

influenza; seasonal variation; COVID-19 pandemic; stringency index

## Introduction

Seasonal influenza is an epidemic disease caused by the influenza virus with a high burden and severity. The World Health Organization (WHO) and the Centers for Disease Control and Prevention of most countries jointly coordinate influenza surveillance and report weekly data on human seasonal influenza viruses, including the activities of A (H1N1), A (H3N2), B/Victoria, and B/Yamagata. Data obtained through virology surveillance in temperate locations have demonstrated a consistent seasonal pattern with this seasonality being weaker in subtropical and tropical locations [1].

The WHO declared the COVID-19 pandemic on March 11, 2020. Accordingly, governments worldwide launched different levels of mitigation strategies. Subsequently, several northern-hemisphere countries showed a notable decrease in the acute respiratory tract infection or influenza-like illness (ILI) consultation rates at the end of the 2019/2020 season and during the 2020/2021 season compared with the rates in previous years [2-4]. In 2020, there were decreased influenza burdens and activity in southern-hemisphere countries, which varied among the countries [5]. In China, compared to the average levels during 2012-2019, the positive rate of influenza virus decreased during 2020 with changes of –87.6% [6]. Similar variations have been observed for other respiratory viruses in China and the United States [6,7].

Decreased influenza transmission dynamics during the 2019/2020 season mainly resulted from the public health and social measures implemented to mitigate the COVID-19 pandemic [8,9]. The reducing outpatient ILI consultations in cities of China were also caused by the mitigation measures [10]. Influenza activity and seasonal patterns were influenced by COVID-19, and the activity of these 2 diseases suppressed and competed with each other, showing a seesaw effect [11]. The study also found that population susceptibility increased in both China and the United States, which might lead to large, high-intensity influenza outbreaks in the 2022/2023 season [12]. Furthermore, the lack of exposure to influenza together with unpredictable antigenic changes may have reduced population immunity, which could lead to a more severe influenza season after the COVID-19 mitigation measures are lifted [13].

Prior to the COVID-19 pandemic, influenza virus subtypes and lineages were cocirculated and the predominant subtypes varied each year, while after the pandemic the B/Victoria lineage dominated in 2020/2021 in Chongqing, China [14]. Large changes occurred in influenza subtype distribution during the pandemic, including the sharp decrease in the global prevalence of B/Yamagata viruses and a temporal increased prevalence of B/Victoria virus in the Western Pacific Region [15]. The unexpected activities of these viruses were potential risks. Therefore, there is a need to elucidate variations in the epidemiology and transmission characteristics of the influenza virus since the COVID-19 pandemic [16]. Meanwhile, it is necessary to monitor and update knowledge on the circulation types and subtypes to prepare for future outbreaks [17]. Moreover, given the uncertainty regarding incoming influenza strains and the influence of COVID-19 prevention measures, it is important to perform surveillance of influenza activity and reevaluate the influence of COVID-19 measures on it. Using influenza surveillance data collected from the WHO, we aimed to describe the epidemiologic characteristics of influenza from 2012 to 2023 as well as to explore the correlation between influenza activity and public health and social measures.

## Methods

### Data Collection

We obtained influenza data from January 1, 2012, to February 5, 2023 (data download on February 17, 2023) for the countries and areas of the United States, and Australia from the publicly available platform WHO FluNet, which is remotely provided by the National Influenza Centers of the Global Influenza Surveillance and Response System (GISRS). We also obtained data for the southern and northern areas of China from the Chinese National Influenza Center database. We analyzed data differently by region in China because influenza epidemiologic features differ between the South and North [8]. Meanwhile, the 2 regions have different climatic characteristics. We chose these areas because they are members of WHO Collaborating Centers, meanwhile, the United States and Australia are also members of the WHO Essential Regulatory Laboratories within GISRS, and we believed that there would be accurate data reported to the GISRS from these areas. Meanwhile, China, the United States, and Australia were selected as countries representing each of the 3 WHO transmission zones, where each zone covers a geographical group of countries with different influenza transmission patterns [18]. China and the United States both include tropical, subtropical, and temperate zones, and they are part of the Eastern Asia influenza transmission zone, and the North America influenza transmission zone, respectively. Australia is mainly temperate and located in the Oceania, Melanesia, and Polynesia influenza transmission zones. The influenza activity from 2012 to 2023 is detailed in Figure 1, Table 1, and Multimedia Appendix 1. Then the influenza-positive rates among surveillance samples are further illustrated in Figure 2, and the proportion of influenza subtypes is detailed in Figure 3.

**Figure 1.** Seasonal influenza activity from 2012 to early 2023 with virus types or subtypes across different areas. Surveillance data are obtained from WHO FluNet and the Chinese National Influenza Center database.

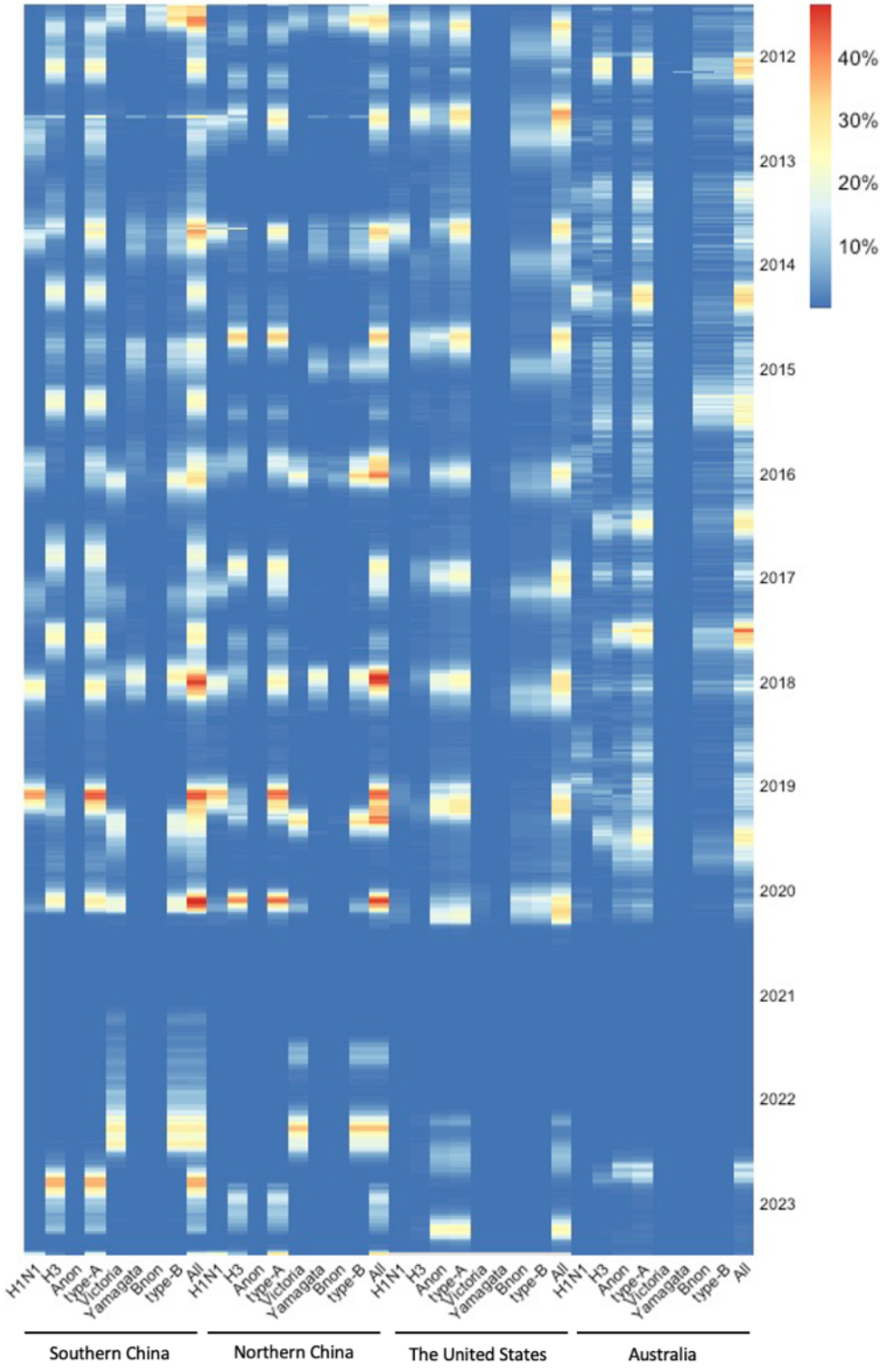


**Table 1.** Comparison of influenza activity features during 2017-2022 in China, the United States, and Australia.

| Years | 2021/2022 | 2020/2021 | 2019/2020 | 2018/2019 | 2017/2018 |
|---|---|---|---|---|---|
| **Southern China**[a] | | | | | |
| Influenza season duration (weeks) | 22 | 0 | 12 | 31 | 39 |
| Influenza activity peak, n/N (%) | 1471/5244 (28.1) | 387/4565 (8.5) | 2262/4676 (48.4) | 4527/9935 (45.6) | 3551/7615 (46.6) |
| Influenza virus type A, n/N (%) | 1/5244 (0) | 0/4565 (0) | 1300/4676 (27.8) | 4503/9935 (45.3) | 1915/7615 (25.2) |
| Influenza virus type B, n/N (%) | 1470/5244 (28) | 387/4565 (8.5) | 962/4676 (20.8) | 24/9935 (0.2) | 1636/7615 (21.5) |
| Week of peak | Week 2, 2022 | Week 39, 2021 | Week 1, 2020 | Week 3, 2019 | Week 3, 2018 |
| **Northern China**[a] | | | | | |
| Influenza season duration (week) | 17 | 0 | 10 | 25 | 18 |
| Influenza activity peak, n/N (%) | 1738/4946 (35.1) | 228/2605 (8.75) | 2363/5031 (47) | 2547/5681 (44.8) | 2874/5931 (48.5) |
| Influenza virus type A, n/N (%) | 0/4946 (0) | 0/2605 (0) | 2175/5031 (43.2) | 2513/5681 (44.2) | 1500/5931 (25.3) |
| Influenza virus type B, n/N (%) | 1738/4946 (35.1) | 228/2605 (8.75) | 188/5031 (3.7) | 34/5681 (0.6) | 1374/5931 (23.2) |
| Week of peak | Week 2, 2022 | Week 21, 2021 | Week 1, 2020 | Week 3, 2019 | Week 1, 2018 |
| **United States**[a] | | | | | |
| Influenza season duration (week) | 0 | 0 | 17 | 19 | 21 |
| Influenza activity peak, n/N (%) | 7682/83,426 (9.2) | 72/26,843 (0.3) | 25,412/77,514 (32.8) | 17,178/58,082 (29.6) | 22,933/75,195 (30.5) |
| Influenza virus type A, n/N (%) | 7652/83,426 (9.2) | 38/26,843 (0.1) | 15,461/77,514 (20) | 16,660/58,082 (28.7) | 18,948/75,195 (25.2) |
| Influenza virus type B, n/N (%) | 30/83,426 (0) | 34/26,843 (0.1) | 9951/77,514 (12.8) | 518/58,082 (0.9) | 3985/75,195 (5.3) |
| Week of peak | Week 15, 2022 | Week 41, 2020 | Week 6, 2020 | Week 9, 2019 | Week 2, 2018 |
| **Australia**[b] | | | | | |
| Influenza season duration (week) | 7 | 0 | 0 | 40 | 12 |
| Influenza activity peak, n/N (%) | 1460/9662 (15.1) | 1/1606 (0.1) | 113/1170 (9.7) | 912/3119 (29.2) | 73/409 (17.9) |
| Influenza virus type A, n/N (%) | 1455/9662 (15.1) | 1/1606 (0.1) | 103/1170 (8.8) | 844/3119 (27.1) | 55/409 (13.5) |
| Influenza virus type B, n/N (%) | 5/9662 (0.1) | 0/1606 (0) | 10/1170 (0.9) | 68/3119 (2.2) | 18/409 (4.4) |
| Week of peak | Week 24, 2022 | Week 6, 2021 | Week 4, 2020 | Week 23, 2019 | Week 7, 2018 |

[a]From week 40 of the previous year to week 39 of the next year.

[b]From week 45 of the previous year to week 44 of the next year.

**Figure 2.** Influenza activity after the COVID-19 pandemic in China, the United States, and Australia. Surveillance data are obtained from WHO FluNet and the Chinese National Influenza Center database. The dark red and red dot lines indicate positivity rates of 10% and 5%, respectively.

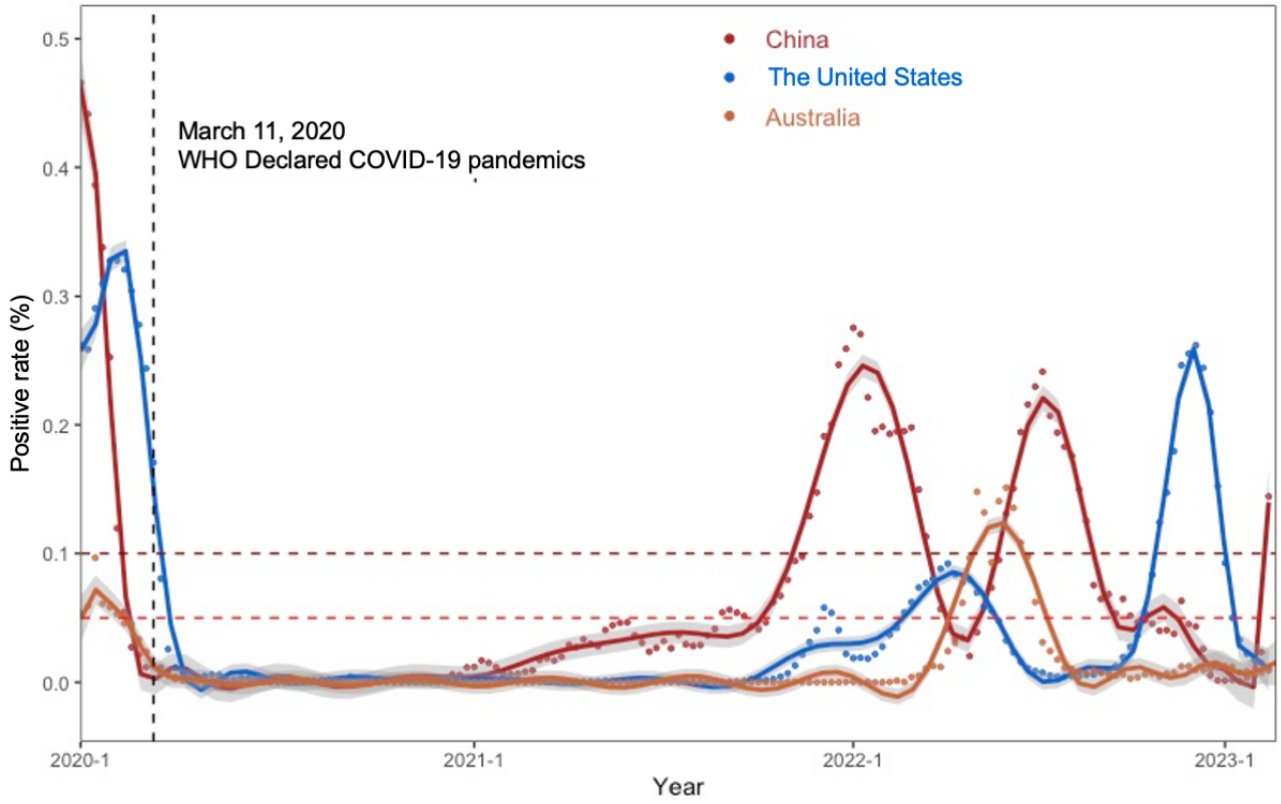
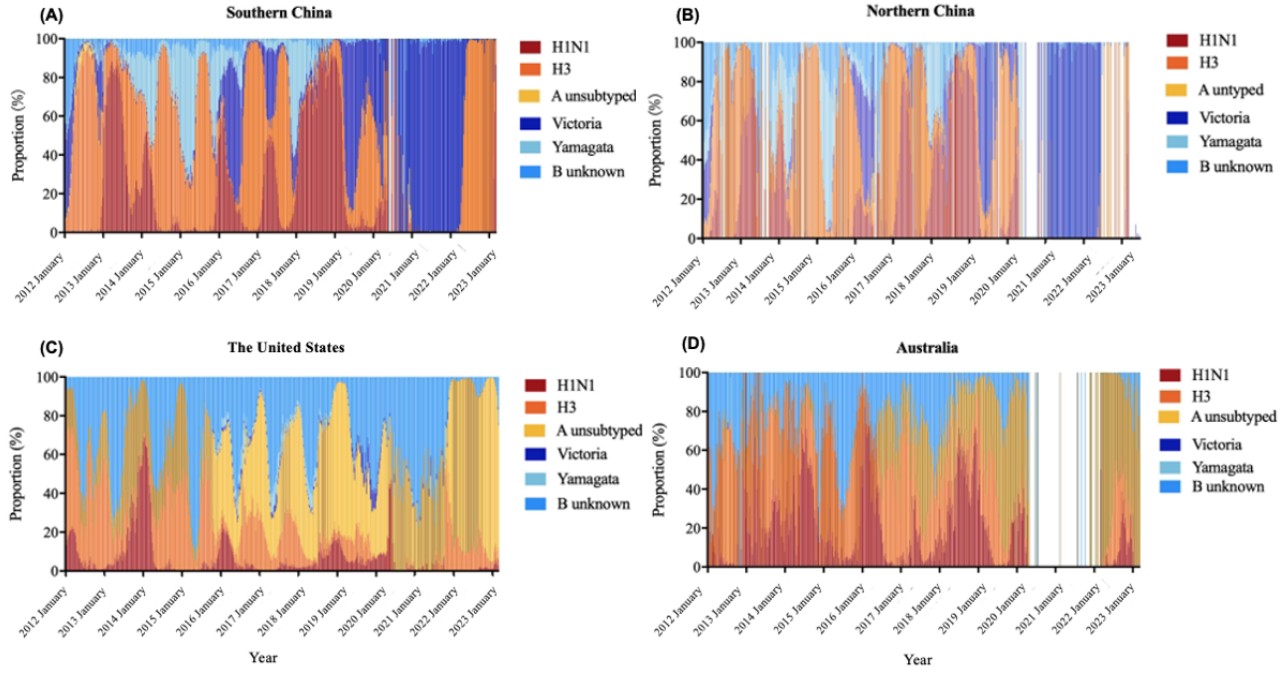



**Figure 3.** Influenza virus subtypes distribution from 2012 to early 2023 among different areas. (A) Southern China. (B) Northern China. (C) The United States. (D) Australia. Red indicates the presence of the H1N1 influenza virus. Orange indicates the H3 influenza virus. Yellow indicates the unsubtyped influenza A. Dark blue indicates the influenza virus of the Victoria lineage. Light blue indicates the influenza virus of the Yamagata lineage. Blue indicates an unknown influenza virus of type B. The white areas indicate no reported cases of influenza from Australia to the FluNet during the period from 2020 week 17 to 2022 week 6, and no reported cases to the Chinese National Influenza Center database from 2020 week 16 to 2020 week 53 in China.



The stringency index (SI) is a composite index for measuring the strictness of "lockdown style" policies that primarily restrict people's behaviors, and it is rescaled to a value from 0 to 100 (100=strictest). It was proposed by the Balavatnik School of Government, University of Oxford [19]. The SI is calculated using 9 metrics as follows: C1, school closing; C2, workplace closing; C3, cancel public events; C4, restrictions on gathering size; C5, close public transport; C6, "shelter-in-place" and home confinement orders; C7, restrictions on internal movements; C8, restrictions on international travel; and H1, public information campaign. Each index for all the 9 strategies was calculated and acquired from the publicly available platform proposed by the Balavatnik School of Government, University of Oxford [20]. This data was downloaded on September 27, 2023, and updated until December 31, 2022. Further analysis of the correlation between influenza activity and SI was detailed in Tables 2 and 3, Figure 4, and Multimedia Appendix 1.

**Table 2.** Influence of individual management measure on influenza virus A and B activity in China, the United States, and Australia during 2020-2022[a].

| Measures | Southern China | | Northern China | | United States | | Australia | |
|---|---|---|---|---|---|---|---|---|
| | $DE_A$[b] (%, *P* value) | $DE_B$ (%, *P* value) | $DE_A$ (%, *P* value) | $DE_B$ (%, *P* value) | $DE_A$ (%, *P* value) | $DE_B$ (%, *P* value) | $DE_A$ (%, *P* value) | $DE_B$ (%, *P* value) |
| C1[c] | 65.8 (<.001)[d] | 43.8 (<.001)[d] | 77.4 (<.001)[d] | 48.2 (<.001)[d] | 88.5 (<.001)[d] | 94 (<.001)[d] | 78 (<.001)[d] | 41.6 (.18) |
| C2[e] | 51 (<.001)[d] | 51.2 (<.001)[d] | 70 (<.001)[d] | 56.5 (<.001)[d] | 80.9 (<.001)[d] | 97.5 (<.001)[d] | 88.5 (<.001)[d] | 69.8 (.04)[d] |
| C3[f] | 38.3 (<.001)[d] | 34.6 (.01)[d] | 59.9 (<.001)[d] | 39.1 (<.001)[d] | 91.1 (<.001)[d] | 90.1 (<.001)[d] | 84.6 (<.001)[d] | 69.4 (.049)[d] |
| C4[g] | 35.9 (<.001)[d] | 36 (<.001)[d] | 55.7 (<.001)[d] | 40.5 (<.001)[d] | 92 (<.001)[d] | 91.2 (<.001)[d] | 87.1 (<.001)[d] | 67.9 (.02)[d] |
| C5[h] | 47.5 (<.001)[d] | 36.4 (.001)[d] | 53.6 (<.001)[d] | 44.7 (<.001)[d] | 83.8 (<.001)[d] | 94.8 (<.001)[d] | 50.3 (<.001)[d] | 73.8 (.06) |
| C6[i] | 37.2 (<.001)[d] | 37.2 (<.001)[d] | 54.2 (<.001)[d] | 54.2 (<.001)[d] | 79.6 (<.001)[d] | 94.9 (<.001)[d] | 66 (<.001)[d] | 43.1 (.19) |
| C7[j] | 36.2 (<.001)[d] | 37.6 (.004)[d] | 51.9 (<.001)[d] | 43.4 (<.001)[d] | 54.1 (<.001)[d] | 77.9 (.001)[d] | 63.2 (<.001)[d] | 62 (.04)[d] |
| C8[k] | 80.7 (<.001)[d] | 53.2 (<.001)[d] | 76.6 (<.001)[d] | 53.2 (<.001)[d] | 65.2 (<.001)[d] | 77.7 (<.001)[d] | 84.3 (<.001)[d] | 46.8 (.03)[d] |
| H1[l] | 37.8 (<.001)[d] | 34.7 (.04)[d] | 59.6 (<.001)[d] | 39.4 (<.001)[d] | 53.9 (<.001)[d] | 97.6 (<.001)[d] | 40.3 (<.001)[d] | 45 (.02)[d] |

[a]A implies the activity of influenza virus type A and B implies the activity of influenza virus type B.

[b]DE: deviance explained.

[c]C1: school closing.

[d]Significant at *P*<.001.

[e]C2: workplace closing.

[f]C3: cancel public events.

[g]C4: restrictions on gathering size.

[h]C5: close public transport.

[i]C6: "shelter-in-place" and home confinement orders.

[j]C7: restrictions on internal movements.

[k]C8: restrictions on international travel.

[k]H1: public information campaign.

**Table 3.** Influence of all the 9 management measures combined on influenza virus A and B activity in China, the United States, and Australia during 2020-2022

| Measures | Southern China | | | Northern China | | | United States | | | Australia | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Deviance explained | *P* value (A) | *P* value (B) | Deviance explained | *P* value (A) | *P* value (B) | Deviance explained | *P* value (A) | *P* value (B) | Deviance explained | *P* value (A) | *P* value (B) |
| | 87.4% (A), 77.6% (B) | | | 83.4% (A), 81.4% (B) | | | 98.6% (A), 99.1% (B) | | | 95.8% (A), 72.7% (B) | | |
| C1[b] | | .72 | <.001[c] | | .003[c] | .04[c] | | .90 | .35 | | .44 | —[d] |
| C2[e] | | .62 | .34 | | .35 | .97 | | .04[c] | .46 | | .06 | .73 |
| C3[f] | | .13 | .48 | | .11 | .16 | | .03[c] | .96 | | .31 | .59 |
| C4[g] | | .15 | .15 | | .22 | .08 | | .22 | .75 | | <.001[c] | .62 |
| C5[h] | | .002[c] | <.001[c] | | .02[c] | <.001[c] | | .01[c] | .76 | | .13 | — |
| C6[i] | | .13 | .001[c] | | .36 | .42 | | .14 | .25 | | .26 | — |
| C7[j] | | .21 | .001[c] | | .16 | .07 | | .61 | .99 | | .43 | .67 |
| C8[k] | | <.001[c] | <.001[c] | | .06 | .001[c] | | .30 | .02[c] | | .04[c] | .80 |
| H1[l] | | .16 | .87 | | .36 | .049 | | .004[c] | <.001[c] | | .30 | .49 |

[a]A implies the activity of influenza virus type A and B implies the activity of influenza virus type B.

[b]C1**:** school closing**.**

[c]Significant at $P<.05$.

[d]Not applicable.

[e]C2: workplace closing.

[f]C3: cancel public events.

[g]C4: restrictions on gathering size.

[h]C5: close public transport.

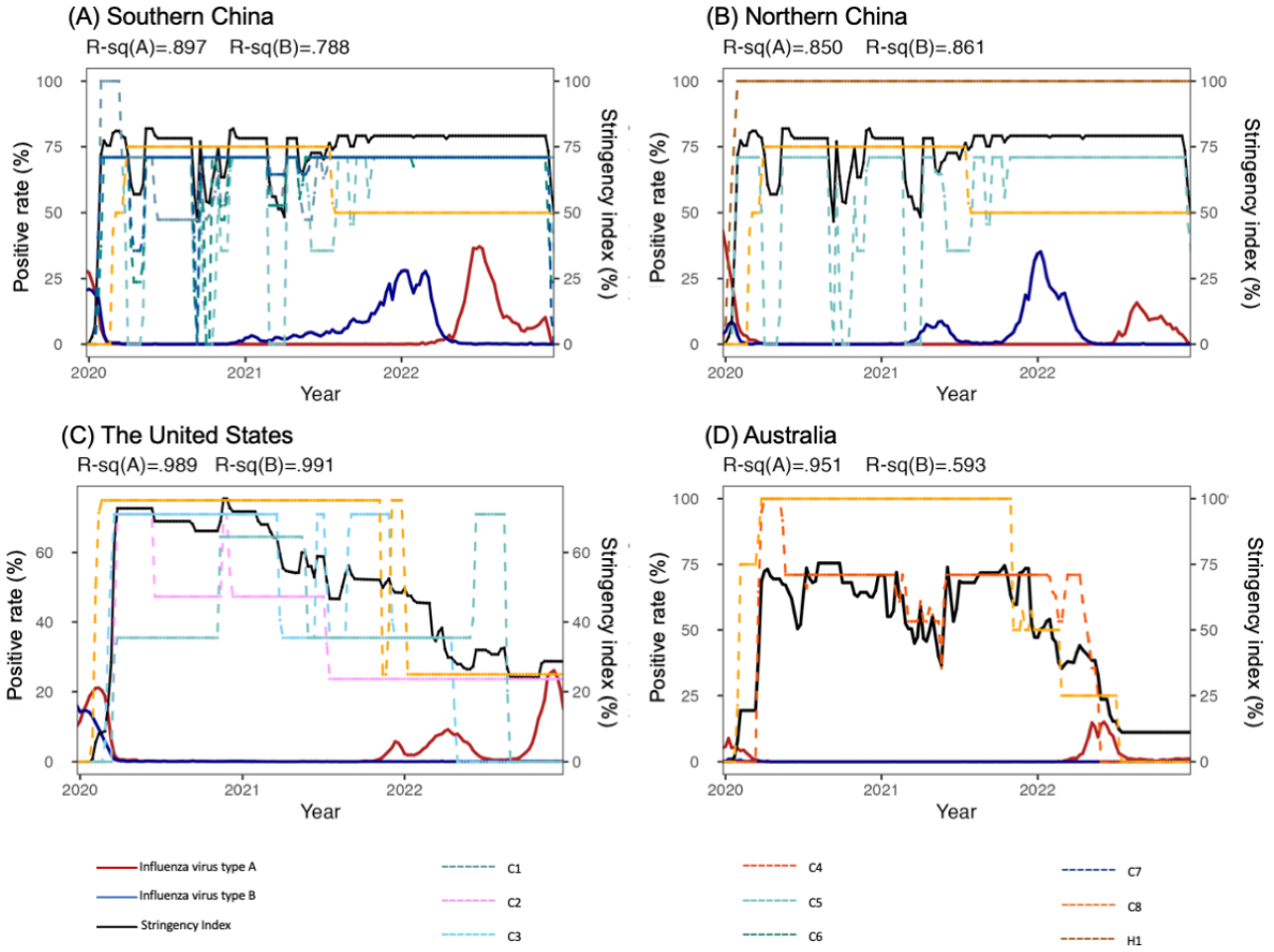
[i]C6: “shelter-in-place” and home confinement orders.

[j]C7: restrictions on internal movements.

[k]C8: restrictions on international travel.

[l]H1: public information campaign.

**Figure 4.** Correlation between the stringency index and the influenza positivity rate in different countries or areas since the COVID-19 pandemic. (A) Southern China. (B) Northern China. (C) The United States. (D) Australia. The dark line represents the stringency index. The red line represents the positivity rate of influenza type A, and the blue line represents the positivity rate of influenza type B. R-sq(A) represents the determination coefficient between influenza type A and the stringency index. In contrast, R-sq(B) represents the determination coefficient between influenza type B and the stringency index. Dashed lines indicate different management measures, including C1, C2, C3, C4, C5, C6, C7, C8, and H1. The measures that significantly affected influenza activity are displayed in the figure. C1: school closing; C2: workplace closing; C3: canceling public events; C4: restrictions on gathering size; C5: closing public transport; C6: stay-at-home requirements; C7: restrictions on internal movement; C8: restrictions on international travel; H1: public information campaign.




## Related Definitions

All of these 3 areas test for influenza virus through ILI cases, and different countries followed their own definition for the diagnosis of ILI. An ILI case in China is defined as a fever (temperature ≥38 ˚C) with a cough or sore throat [21]. ILI in the United States is defined as a fever (temperature of 100 °F [37.8 °C] or greater) and a cough or a sore throat [22]. The Communicable Diseases Network Australia guidelines define ILI as a triad of fever ≥38 °C, respiratory symptoms, and systemic symptoms [23]. The WHO case definition for ILI is an acute respiratory tract infection with temperature ≥38 ˚C and cough, with onset within the last 10 days [23].

The number of specimens that tested positive for influenza virus (with confirmation of the type and subtype) is presented as a percentage of the total number of tested specimens. An influenza positivity rate of ≥10% and <10% for ≥2 consecutive weeks indicates the start and end of seasonal influenza activity, respectively. Weeks were defined using the International Organization for Standardization 8601 standard weeks. Each season was analyzed from week 40 of the previous year to week 39 of the next year in China and the United States, and for Australia, it was from the 45th week of the previous year to the 44th week of the next year.

## Statistical Analysis

Heat maps were applied to depict the influenza virus subtypes epidemiologic characteristics from 2012 to early 2023 (Figure 1). The influenza virus subtypes proportion was illustrated from 2012 to early 2023, and graphs were generated by Prime 8 (version 8.3.1; GraphPad Software, LLC; Figure 3). In the current analysis, we used the SI as an index for the evaluation of public health and social measures (Tables 2 and 3, Figure 4, and Multimedia Appendix 1). The value of the index is the average of 9 subindices about the individual policy indicators, each taking a value from 0 to 100. The Balavatnik School of Government provided daily SI, and the daily SI data were converted to weekly averages to be consistent with the ILI data, detailed in Multimedia Appendix 1. The general additive model was used to assess the correlation between the SIs (defined as an independent variable, $X_i$) and the influenza-positive rate

(defined as the dependent variable, $Y_t$). The deviance explained was calculated to estimate the contribution value of the independent variable to the dependent variable, and the determination coefficient ($R^2$) was calculated to assess the goodness of fit of the models (Multimedia Appendix 1).

Statistical analyses were conducted using R (version 4.2.3; The R Foundation). Statistical significance was set at $P<.05$.

### Ethical Considerations

This study was conducted with data from publicly available platforms, where the data were published. The research involves the use of collections of information or data from which all personal identifiers have been removed before being received by the researchers. Additionally, according to the ethical requirement from China (article 32), the United States (page 4), and Australia (section 5, article 5.1.17), the research using these data may be eligible for a grant of exemption from ethics review [24-26].

## Results

### Characteristics Regarding Seasonal Influenza Activity From 2012 to Early 2023

We used over 200,000 influenza surveillance data in the current analysis. In mainland China, the United States, and Australia, the influenza activity showed obvious seasonal peaks previous to the COVID-19 pandemic. Figure 1 shows the distribution of all influenza genotypes from 2012 to 2023 (Figure 1). The influenza epidemiologic features in these 4 regions differed before and during the pandemic, and the new epidemiologic features are also different in each area. Since the COVID-19 pandemic, there has been a gradual decrease in the influenza positivity rate, which was <10% in China and Australia by the time WHO declared COVID-19 a pandemic (Figure 2). The influenza positivity rate remained <5% during the 2020/2021 season in all 4 regions. Influenza activity remained low in the United States, and Australia during the 2021/2022 season, with peak positivity rates of 9.2% (mainly influenza A), 7682 of 83,426 surveillance samples, in week 15, 2022, and 15.1% (mainly influenza A), 1460 of 9662 surveillance samples, in week 24, 2022, respectively. During the 2021/2022 season, influenza activity significantly increased compared with the previous year in southern and northern China, with peak positivity rates of 28.1% (mainly influenza B), 1471 of 5244 surveillance samples, and 35.1% (mainly influenza B), 1738 of 4946 surveillance samples, in the second week of 2022, respectively, and with 2 seasonal influenza activity peaks during winter and summer in northern instead of southern China (Figure 2 and Table 1).

For China, the influenza season durations were significantly shorter during the 2019/2020, and 2020/2021 seasons, however, the durations tended to be similar to the prepandemic during the 2021/2022 season (22 weeks in the south and 17 weeks in the north). The influenza season durations were significantly shorter during the 2020/2021 (0 weeks) and 2021/2022 seasons (0 weeks) in the United States, and during the 2019/2020 (0 weeks), 2020/2021 (0 weeks), and 2021/2022 seasons (7 weeks) in Australia (Table 1). For China and Australia, compared to the 2020/2021 season, the influenza season duration was much longer, and the activity peaks were higher during the 2021/2022 season (Table 1). For the United States, influenza activity increased with a positivity rate of over 10% at the end of 2022 (Figure 2).

### Variations in the Features of the Influenza Virus Since the COVID-19 Pandemic

Figure 3 details the influenza virus types and subtypes. During the last 8 prepandemic years, type A was the dominant influenza virus genotype for most of the time in all 3 areas. The dominant influenza virus genotype after the COVID-19 pandemic gradually became type B (mainly Victoria) in China during the winter of 2021/2022 (proportion >98%, 24,541/24,908 in the South and 20,543/20,634 in the North). Then in April 2022, type B/Victoria was decreased. However, type A (H3) became the dominant subtype (proportion >99%) in southern China (18,667/18,822) first and in northern China (3268/3278) later. Contrastingly, the dominant influenza virus genotype during the 2021/2022 season was type A in the United States and Australia (proportion >90%, 153,735/155,565 and 8752/8769, respectively) for the 2 countries. In all these 4 areas, influenza virus type B/Yamagata has not been detected since the COVID-19 pandemic.

### Correlation of Public Health and Social Measures for COVID-19 With Seasonal Influenza Activity

Seasonal influenza activity was impacted by the COVID-19 pandemic. We analyzed the correlation between the SI and the influenza positivity rate (Tables 2 and 3, and Figure 4). All 9 management measures were individually evaluated and detailed in Table 2. Most of these measures significantly affected the positive rate of influenza types A and B. However, for influenza B in Australia, measures C1 (school closing), C5 (closing public transport), and C6 (shelter-in-place and home confinement orders) affected the positive rate without significant differences (Table 2).

We further analyzed the combined effect of all measures on influenza-positive rates (Table 3 and Figure 4). When combined, measures C1 (school closing), C5 (closing public transport), C6 (shelter-in-place and home confinement orders), C7 (restrictions on internal movements), and C8 (restrictions on international travel) significantly affected influenza A or B activity ($P<.05$) in southern China. In northern China, C1 and C5 significantly affected influenza (both A and B) activity ($P<.05$), and C8 significantly affected influenza B activity ($P<.05$). The deviance explained value for influenza A was 87.4% and for influenza B was 77.6% in southern China, while for influenza A and B in northern China, it was 83.4% and 81.4%, respectively. In the United States, C2 (workplace closing), C3 (cancel public events), C5, and H1 (public information campaign) significantly affected the influenza A positive rate, and C8 and H1 affected the positive rate of influenza B. The correlation between the SI and influenza activity was relatively stronger, with deviance-explained values of 98.6% for influenza A and 99.1% for influenza B in the United States (Table 3). In Australia, C5 and C8 affected influenza A activity significantly ($P<.05$). As the positive rate of influenza B was extremely low from 2020 to 2022, no

significant effects were found on the influenza B activity. The correlation between the SI and the influenza positivity rate in Australia was 95.8% for influenza type A and 72.7% for influenza type B. However, despite the high level of the SI in China, influenza activity increased in 2021 compared to 2020.

## Discussion

### Principal Results

We analyzed influenza surveillance data from China, the United States, and Australia to describe influenza activity within 10 years before and after the COVID-19 pandemic. The implementation of public health and social measures in China and the United States contributed to decreased influenza activity during the 2019/2021 season [7,8]. However, we observed a gradual increase in influenza activity starting from the beginning of 2021 reaching 28.1% (1471/5244) in southern China and 35.1% (1738/4946) in northern China by the second week of 2022. Contrastingly, influenza activity in the United States, and Australia during the 2021/2022 season remained low, even though it was slightly higher than during the 2020/2021 season (Figure 2 and Table 1). However, by the end of 2022 and early 2023, influenza activity increased in the United States, with a positivity rate of 26.2% in the 49th week of 2022. Notably, the dominant influenza virus genotype since the COVID-19 pandemic in China, especially during the 2021/2022 season, was type B/Victoria, which then transformed to type A (H3) in April 2022. The dominant virus was type A in the United States and Australia during 2022 and early 2023. Postpandemic variations in influenza activity were correlated with the public health and social measures for COVID-19. However, it is important to consider that there may be other factors affecting influenza activity and emerging as new epidemiological features.

In Canada, surveillance data obtained during the 2021/2022 season revealed persistent sporadic influenza activity, with the predominant virus subtype being A (H3N2) [27]. Similarly, type A (H3) was predominant in the United States, with laboratory-confirmed influenza showing a low positivity rate during the 2021/2022 season. However, the positivity rate increased at the end of 2022 and then decreased at the beginning of 2023. In Australia, type A (H3) was dominant in May and June 2022, and then the influenza positivity decreased and maintained a low level until 2023. China showed unique changes in the dominant influenza virus type, which was type B/Victoria during the 2021/2022 season. Another study found that before the COVID-19 outbreak, influenza A viruses dominated in China, whereas influenza B/Victoria viruses dominated during the 2020/2021 season [15]. Our results underscored that type B/Victoria dominated in China until April 2022, with a positivity rate of over 25% in both southern and northern regions during the 2021/2022 winter. During the influenza season, type B usually results in mild symptoms and less frequent epidemics and has never caused pandemics in humans [28]. The epidemiological characteristics of influenza during the 2021/2022 season in China, specifically before January 2022, included single-type spreading and a relatively higher activity peak compared to the previous year. These characteristics could be attributed to restricted inbound travels and within-country transmission, which was also the dominant transmission mechanism for COVID-19 in China [29]. As results displayed in our study, in China during 2020-2021, the SI remained at a high level. However, the restrictions on population internal movements were fluctuant, and sometimes even measures such as “restrictions on internal movements” or “close public transport” were canceled (Figure S1 in Multimedia Appendix 1). This led to the emergence of influenza cases in 2021, albeit at a low level. Compared to domestic movement mitigation, the mitigation of international travel had a larger effect [30]. The measure of restrictions on international travel significantly affected the activities of influenza in China, and the positive rate of influenza (B/Victoria) increased by over 10% after the relaxation of international travel in 2021 (Figure 4 and Table 3). Influenza type A (H3N2) first emerged in southwestern China in early 2022, extensively spreading in southern China, and finally reaching northern China in July 2022. It may have entered China from Southeast Asia.

Currently, promoting vaccination coverage remains the primary strategy for controlling influenza [31]. The COVID-19 pandemic increased the likelihood of receiving an influenza vaccine in the United States [32], England [13], and Australia [33]. However, in China, the influenza vaccination rate remains low; moreover, parental acceptance of childhood vaccination is still low (29% in northwestern China) [34]. This could explain the relatively higher influenza activity in China compared with the United States, and Australia.

The low influenza activity during the 2020/2021 season is mainly attributable to public health and social measures [8]. A modeling study predicted that influenza activity would return to prepandemic levels during the 2021/2022 season after easing the current mitigation measures for COVID-19 [35]. Even though significant correlations between the SI and influenza positivity rates were found in Australia and the United States, these correlations were mainly driven by the early phase of the pandemic and fluctuations in mitigation measures during 2022. Although we observed a relatively higher influenza activity peak during the 2021/2022 season in China, the local mitigation measures were still strictly implemented, as indicated by the SI. The potential explanation might be that the SI was a composite index, and the high number of SIs for China might be driven by international travel controls instead of restrictions on internal movements (Figure S1 in Multimedia Appendix 1). However, there are differences in the effectiveness of various mitigation measures in reducing virus transmission. For example, a previous study showed that mask-wearing was more effective than mitigating travel or movement, while interventions targeting international travel were more effective than those targeting domestic travel [30]. Implementing optimized strategies with composite measures would help in quickly responding to unpredictable respiratory infectious diseases. Furthermore, it is necessary to evaluate the effectiveness of different measures, which can help establish optimized strategies in the future.

There are immunopathological similarities between the influenza virus and SARS-CoV-2 [36]. Additionally, influenza vaccination is associated with reduced susceptibility to or severity of COVID-19 [37]. Therefore, the observed correlations in this

study might not be limited to the influence of public health and social measures. Multi-pathogen mutual interference might also play an important role in current influenza activities. During the current unstable phase, it is important to monitor influenza and other respiratory pathogens' activities simultaneously. Rapid integration of multi-pathogen testing and surveillance would help prevent emerging pathogens from evolving into another pandemic [38]. Currently, it is crucial and timely to track the activity of influenza and re-evaluate the effect of COVID-19 mitigation measures, because as the pandemic progresses and after, more factors will emerge that affect influenza activity. Meanwhile, it is also essential to strictly surveil the seasonal influenza activity and the variation of pathogens because the COVID-19 pandemic influenced the features of seasonal activities.

### Limitations

This study has several limitations. First, the evaluation of the effect of other measures, such as mask-wearing, was not supported by the currently available data, despite their reported significant impact on influenza activities. Second, we did not consider climate and environmental factors. In the early stage of the COVID-19 pandemic, public health and social measures had the strongest influence on influenza activity [6]. However, temperature, wind speed, and a particulate matter level of 2.5 also influence the transmission of some respiratory infectious diseases [39].

### Conclusions

This study provides an updated report on influenza activity in 4 areas from 2012 to early 2023, as well as the impact of public health and social measures for COVID-19 on influenza activity. Our findings indicate an increase in influenza activity in these 3 countries. While there were correlations between mitigation measures for COVID-19 and influenza activity, the main driving force was the early stage of the pandemic. The role of mitigation management in influencing influenza activity varied across different stages of the pandemic and among different countries. It is crucial to maintain vigilance to prevent influenza and other respiratory pathogens from causing epidemics. Further studies are warranted to investigate the reasons behind the variations in dominant virus subtypes across regions and to assess the influence of individual mitigation measures, climate factors, and the activity of other pathogens on influenza.

## Acknowledgments

We thank all the members who collected and submitted the influenza surveillance data of China, the United States, and Australia to the WHO FluNet. Additionally, we also thank all the members working on calculating and providing the SI.

## Data Availability

Data on influenza activity are available on the WHO FluNet and data on the SI are available on the website of the Balavatnik School of Government, the University of Oxford. Processed datasets used in this analysis are available from the corresponding author upon reasonable request.

## Authors' Contributions

MJiang provided the research concept, collected the data, analyzed the data, and drafted this paper. MJia commented on and revised this paper. YS collected the data, commented on, and revised this paper. WY provided the research concept, commented on, and revised this paper.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Material, tables, and figures for the methods and results.
[DOCX File , 384 KB - ijmr_v13i1e47370_app1.docx ]

## References

1. Neumann G, Kawaoka Y. Seasonality of influenza and other respiratory viruses. EMBO Mol Med 2022 Apr 07;14(4):e15352 [FREE Full text] [doi: 10.15252/emmm.202115352] [Medline: 35157360]
2. Adlhoch C, Sneiderman M, Martinuka O, Melidou A, Bundle N, Fielding J, European Influenza Surveillance Network, Members of the European Influenza Surveillance Network. Spotlight influenza: The 2019/20 influenza season and the impact of COVID-19 on influenza surveillance in the WHO European Region. Euro Surveill 2021 Oct;26(40):2100077 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.40.2100077] [Medline: 34622760]
3. Li Q, Wang J, Lv H, Lu H. Impact of China's COVID-19 prevention and control efforts on outbreaks of influenza. Biosci Trends 2021 Jul 06;15(3):192-195 [FREE Full text] [doi: 10.5582/bst.2021.01242] [Medline: 34176827]
4. Uyeki TM, Wentworth DE, Jernigan DB. Influenza activity in the US during the 2020-2021 season. JAMA 2021 Jun 08;325(22):2247-2248. [doi: 10.1001/jama.2021.6125] [Medline: 34028492]

5. Kim J, Gómez Gómez RE, Hong K, Yum S, Jang J, Chun BC. Changing influenza activity in the Southern hemisphere countries during the COVID-19 pandemic. Int J Infect Dis 2021 Jul;108:109-111 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.039] [Medline: 34022335]
6. Li ZJ, Yu LJ, Zhang HY, Shan CX, Lu QB, Zhang XA, Chinese Centers for Disease ControlPrevention (CDC) Etiology Surveillance Study Team of Acute Respiratory Infections. Broad impacts of coronavirus disease 2019 (COVID-19) pandemic on acute respiratory infections in China: an observational study. Clin Infect Dis 2022 Aug 24;75(1):e1054-e1062 [FREE Full text] [doi: 10.1093/cid/ciab942] [Medline: 34788811]
7. Olsen SJ, Winn AK, Budd AP, Prill MM, Steel J, Midgley CM, et al. Changes in influenza and other respiratory virus activity during the COVID-19 pandemic—United States, 2020-2021. MMWR Morb Mortal Wkly Rep 2021 Jul 23;70(29):1013-1019 [FREE Full text] [doi: 10.15585/mmwr.mm7029a1] [Medline: 34292924]
8. Feng L, Zhang T, Wang Q, Xie Y, Peng Z, Zheng J, et al. Impact of COVID-19 outbreaks and interventions on influenza in China and the United States. Nat Commun 2021 May 31;12(1):3249 [FREE Full text] [doi: 10.1038/s41467-021-23440-1] [Medline: 34059675]
9. Lee H, Lee H, Song K, Kim E, Park J, Jung J, et al. Impact of public health interventions on seasonal influenza activity during the COVID-19 outbreak in Korea. Clin Infect Dis 2021 Jul 01;73(1):e132-e140 [FREE Full text] [doi: 10.1093/cid/ciaa672] [Medline: 32472687]
10. Zhang X, Du J, Li G, Chen T, Yang J, Yang J, et al. Assessing the impact of COVID-19 interventions on influenza-like illness in Beijing and Hong Kong: an observational and modeling study. Infect Dis Poverty 2023 Feb 16;12(1):11 [FREE Full text] [doi: 10.1186/s40249-023-01061-8] [Medline: 36797765]
11. Wang Q, Jia M, Jiang M, Liu W, Yang J, Dai P, et al. Seesaw effect between COVID-19 and influenza from 2020 to 2023 in World Health Organization regions: correlation analysis. JMIR Public Health Surveill 2023 Jun 12;9:e44970 [FREE Full text] [doi: 10.2196/44970] [Medline: 37191650]
12. Wang Q, Jia M, Jiang M, Cao Y, Dai P, Yang J, et al. Increased population susceptibility to seasonal influenza during the COVID-19 pandemic in China and the United States. J Med Virol 2023 Oct;95(10):e29186. [doi: 10.1002/jmv.29186] [Medline: 37855656]
13. Ashraf MA, Rajaram S, English PM. How the COVID 19 pandemic will shape influenza public health initiatives: the UK experience. Hum Vaccin Immunother 2022 Nov 30;18(5):2056399 [FREE Full text] [doi: 10.1080/21645515.2022.2056399] [Medline: 35435806]
14. Li Z, Xiong Y, Long J, Li T, Fu X, Yang S, et al. Resurgence of influenza during COVID-19 in Chongqing, China: a retrospective analysis. J Med Virol 2023 Nov;95(11):e29249. [doi: 10.1002/jmv.29249] [Medline: 38009822]
15. Zheng L, Qi J, Wu J, Zheng M. Changes in influenza activity and circulating subtypes during the COVID-19 outbreak in China. Front Med (Lausanne) 2022;9:829799 [FREE Full text] [doi: 10.3389/fmed.2022.829799] [Medline: 35391894]
16. Ryu S, Cowling BJ. Human influenza epidemiology. Cold Spring Harb Perspect Med 2021 Dec 01;11(12):a038356 [FREE Full text] [doi: 10.1101/cshperspect.a038356] [Medline: 32988982]
17. Zheng L, Lin Y, Yang J, Fang K, Wu J, Zheng M. Global variability of influenza activity and virus subtype circulation from 2011 to 2023. BMJ Open Respir Res 2023 Jul;10(1):e001638 [FREE Full text] [doi: 10.1136/bmjresp-2023-001638] [Medline: 37491131]
18. Influenza transmission zones. WHO. 2018. URL: https://www.who.int/publications/m/item/influenza_transmission_zones [accessed 2024-09-14]
19. COVID-19: stringency index. Oxford Uo. URL: https://tinyurl.com/bd66sz5w [accessed 2024-09-14]
20. OxCGRT covid-policy-dataset. Blavatnik School of Government UoO. 2020. URL: https://github.com/OxCGRT/covid-policy-dataset [accessed 2024-10-01]
21. Center C. Technical guidelines for national influenza surveillance. Center CNI. 2017. URL: https://ivdc.chinacdc.cn/cnic/zyzx/jcfa/201709/P020170930331067634607.pdf [accessed 2024-09-14]
22. Influenza surveillance: purpose and methods. Prevention cfDCa. URL: https://www.cdc.gov/fluview/overview/?CDC_AAref_Val=https://www.cdc.gov/flu/weekly/overview.htm [accessed 2023-10-13]
23. Influenza infection (flu) - CDNA national guidelines for public health units. Australian Government DoHaAC. URL: https://www.health.gov.au/resources/publications/influenza-infection-flu-cdna-national-guidelines-for-public-health-units?language=en [accessed 2019-01-18]
24. Notice on the issuance of the ethical review methods for life sciences and medical research involving human beings. Education DoSaT. URL: http://app.gcph.com.cn/upload/files/2023/2/2814189268.pdf [accessed 2024-09-14]
25. Coded private information or specimens use in research, guidance (2008). Services USDoHaH. 2008. URL: https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html [accessed 2024-09-14]
26. National Statement on Ethical Conduct in Human Research 2023. Australian Government NHMRC. URL: https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2023 [accessed 2024-09-14]
27. Bancej C, Rahal A, Lee L, Buckrell S, Schmidt K, Bastien N. National FluWatch mid-season report, 2021-2022: sporadic influenza activity returns. Can Commun Dis Rep 2022 Jan 26;48(1):39-45 [FREE Full text] [doi: 10.14745/ccdr.v48i01a06] [Medline: 35273468]

28. Javanian M, Barary M, Ghebrehewet S, Koppolu V, Vasigala V, Ebrahimpour S. A brief review of influenza virus infection. J Med Virol 2021 Aug;93(8):4638-4646. [doi: 10.1002/jmv.26990] [Medline: 33792930]
29. Han X, Xu Y, Fan L, Huang Y, Xu M, Gao S. Quantifying COVID-19 importation risk in a dynamic network of domestic cities and international countries. Proc Natl Acad Sci U S A 2021 Aug 03;118(31):e2100201118 [FREE Full text] [doi: 10.1073/pnas.2100201118] [Medline: 34285082]
30. Han S, Zhang T, Lyu Y, Lai S, Dai P, Zheng J, et al. Influenza's plummeting during the COVID-19 pandemic: the roles of mask-wearing, mobility change, and SARS-CoV-2 interference. Engineering (Beijing) 2023 Feb;21:195-202 [FREE Full text] [doi: 10.1016/j.eng.2021.12.011] [Medline: 35127196]
31. Gao G, Liu W. Let's get vaccinated for both flu and COVID-19: on the World Flu Day 2021. China CDC Wkly 2021 Oct 29;3(44):915-917 [FREE Full text] [doi: 10.46234/ccdcw2021.227] [Medline: 34745691]
32. Li K, Yu T, Seabury SA, Dor A. Trends and disparities in the utilization of influenza vaccines among commercially insured US adults during the COVID-19 pandemic. Vaccine 2022 Apr 26;40(19):2696-2704 [FREE Full text] [doi: 10.1016/j.vaccine.2022.03.058] [Medline: 35370018]
33. Del Riccio M, Lina B, Caini S, Staadegaard L, Wiegersma S, Kynčl J, et al. Letter to the editor: increase of influenza vaccination coverage rates during the COVID-19 pandemic and implications for the upcoming influenza season in northern hemisphere countries and Australia. Euro Surveill 2021 Dec;26(50):2101143 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.50.2101143] [Medline: 34915972]
34. Lu X, Wang J, Hu L, Li B, Lu Y. Association between adult vaccine hesitancy and parental acceptance of childhood COVID-19 vaccines: a web-based survey in a northwestern region in China. Vaccines (Basel) 2021 Sep 27;9(10):1088 [FREE Full text] [doi: 10.3390/vaccines9101088] [Medline: 34696196]
35. Han S, Zhang T, Lyu Y, Lai S, Dai P, Zheng J, et al. The incoming influenza season—China, the United Kingdom, and the United States, 2021-2022. China CDC Wkly 2021 Dec 03;3(49):1039-1045 [FREE Full text] [doi: 10.46234/ccdcw2021.253] [Medline: 34934512]
36. Khorramdelazad H, Kazemi MH, Najafi A, Keykhaee M, Zolfaghari Emameh R, Falak R. Immunopathological similarities between COVID-19 and influenza: investigating the consequences of co-infection. Microb Pathog 2021 Mar;152:104554 [FREE Full text] [doi: 10.1016/j.micpath.2020.104554] [Medline: 33157216]
37. Su W, Wang H, Sun C, Li N, Guo X, Song Q, et al. The association between previous influenza vaccination and COVID-19 infection risk and severity: a systematic review and meta-analysis. Am J Prev Med 2022 Jul;63(1):121-130 [FREE Full text] [doi: 10.1016/j.amepre.2022.02.008] [Medline: 35410774]
38. Li C, Zhong H, Xie Y, Bai T, Yan B, Sonne C. Speed up multi-pathogen surveillance. Lancet 2023 Feb 04;401(10374):345-346 [FREE Full text] [doi: 10.1016/S0140-6736(23)00134-4] [Medline: 36739134]
39. Deji Z, Tong Y, Huang H, Zhang Z, Fang M, Crabbe MJC, et al. 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 Mar 25;13:e43585 [FREE Full text] [doi: 10.2196/43585] [Medline: 38526532]


## Abbreviations

**GISRS:** Global Influenza Surveillance and Response System
**ILI:** influenza-like illness
**SI:** stringency index
**WHO:** World Health Organization


*Edited by T de Azevedo Cardoso; submitted 17.03.23; peer-reviewed by I Ahn, K Ferrini, D Dinh; comments to author 22.09.23; revised version received 28.10.23; accepted 28.08.24; published 09.10.24.*

*Please cite as:*
*Jiang M, Jia M, Wang Q, Sun Y, Xu Y, Dai P, Yang W, Feng L*
*Changes in the Epidemiological Features of Influenza After the COVID-19 Pandemic in China, the United States, and Australia: Updated Surveillance Data for Influenza Activity*
*Interact J Med Res 2024;13:e47370*
*URL: https://www.i-jmr.org/2024/1/e47370*
*doi:10.2196/47370*
*PMID:39382955*

Original Paper

# Dynamics of Blood Lipids Before, During, and After Diurnal Fasting in Inactive Men: Quasi-Experimental Study

Khalid Aljaloud[1], PhD; Naif Al-Barha[2,3], PhD; Abeer Noman[4], PhD; Abdulaziz Aldayel[1], PhD; Yahya Alsharif[1], PhD; Ghareeb Alshuwaier[1], PhD

[1]Department of Exercise Physiology, College of Sport Sciences and Physical Activity, King Saud University, Riyadh, Saudi Arabia

[2]Department of Physical Activity, Faculty of Education and Science, Albaydha University, Rada'a, Yemen

[3]School of Physical Education and Sport Training, Shanghai University of Sport, Shanghai, China

[4]Department of Food Science and Technology, Faculty of Agriculture, Sana'a University, Sana'a, Yemen

**Corresponding Author:**
Khalid Aljaloud, PhD
Department of Exercise Physiology
College of Sport Sciences and Physical Activity
King Saud University
P O Box 2454
Riyadh, 11451
Saudi Arabia
Phone: 966 0118063100
Email: khaljaloud@ksu.edu.sa

## Abstract

**Background:** There is a lack of investigation into the dynamics of blood lipids before, during, and after diurnal fasting, especially in inactive men.

**Objective:** This study determined dynamic changes in blood lipids in inactive men before, during, and after they underwent diurnal fasting.

**Methods:** A total of 44 young men aged a mean 27.6 (SD 5.8) years were recruited to evaluate their habitual physical activity and diet using a questionnaire developed for this study. Body composition was evaluated using a bioelectrical impedance analysis machine (Tanita BC-980). An 8-ml blood sample was collected to evaluate blood lipids and glucose. All measurements were taken 2-3 days before Ramadan, during Ramadan (at week 2 and week 3), and 1 month after Ramadan. A 1-way repeated measures ANOVA was used to compare the measured variables before, during, and after the month of Ramadan. When a significant difference was found, post hoc testing was used. Differences were considered significant at $P<.05$.

**Results:** There was a significant reduction in low-density lipoprotein during Ramadan compared to before and after Ramadan (83.49 mg/dl at week 3 vs 93.11 mg/dl before Ramadan [$P=.02$] and 101.59 mg/dl after Ramadan [$P=.007$]). There were significant elevations in fasting blood glucose (74.60 mmol/L before Ramadan vs 81.52 mmol/L at week 3 [$P=.03$] and 86.51 mmol/L after Ramadan [$P=.01$]) and blood pressure (109 mm Hg before Ramadan vs 114 mm Hg after Ramadan; $P=.02$) reported during and even after the month of Ramadan, although both fasting blood glucose and blood pressure were within normal levels.

**Conclusions:** Ramadan fasting could be an independent factor in reducing low-density lipoprotein. Further investigations are encouraged to clarify the impact of diurnal fasting on blood lipids in people with special conditions.

***(Interact J Med Res 2024;13:e56207)*** doi:10.2196/56207

**KEYWORDS**

cardiovascular diseases; cardiovascular risk factors; lipids; glucose measurement; fasting; Ramadan; body composition

## Introduction

Ramadan is a lunar month that varies between 29 and 30 days. The world's Muslim population is about 1.5 billion [1] and is expected to increase to 2.2 billion by 2030 [2]. Ramadan fasting is a form of religious diurnal fasting in which adult Muslims abstain from having food and drink (unless necessary) during the daylight hours, from dawn to sunset. The average diurnal fasting time during Ramadan month is 15 (SD 3) hours per day, depending on the season and the country's geographical location.

During the month of Ramadan, Muslims usually change their daily lifestyle [3,4]. Muslim adults usually become less active during Ramadan, especially in the daytime compared to the nighttime [5-7]. These changes, in addition to diurnal fasting, may negatively affect health-related blood biomarkers. Previous evidence concluded that metabolic disease biomarkers, including blood lipids, were elevated in women and urban residents [6]. Genetic factors are key in metabolic disease biomarkers, including high blood cholesterol, in different ethnic groups [8]. For example, some studies reported heritability of up to 70% for genes involving high-density lipoprotein (HDL) [9]. Diurnal fasting has been found to improve blood lipids [4,5,10,11].

The negative role of physical inactivity on blood lipids is evident. Physical inactivity has been found to play a key role in developing cardiovascular and metabolic disease risk factors such as obesity [12-15]. In a recent study conducted with young Saudi medical students, one of the most common risk factors of cardiovascular and metabolic disease was physical inactivity (57.9%) [16]. Insufficient physical activity is responsible for around 10% of all deaths globally, and sedentary behavior (SB) costs global health care organizations billions of dollars each year. Furthermore, the accumulation of low-density lipoprotein (LDL) cholesterol in artery walls causes narrowing of the arteries, which may lead to the development of cardiovascular disease (CVD) [17].

The number of studies investigating the association between Ramadan fasting and health promotion dramatically increased from 2010 to 2021, with 1276 studies being published [18]. Although there is robust available evidence exploring the impact of Ramadan fasting on health-related biomarkers, reports on the effect of diurnal fasting on blood lipids in inactive adults are scattered [5,11,18-20]. The impact of fasting during Ramadan on body composition and physiological parameters has been widely investigated [21-24]. During the month of Ramadan, people may achieve a significant reduction in body fat mass and blood lipids such as LDL. Some studies did not report significant improvement in triglyceride (TG) levels and fasting blood glucose (FBG) [23]. On the other hand, some blood and metabolic changes that occur during Ramadan fasting are temporary [21,25,26]. It seems that the positive change in LDL levels as a result of diurnal Ramadan fasting is one of these temporary changes [25]. The major changes in blood parameters observed in previous studies could have been due to the lifestyle of participants [22,24].

Despite the results of previous studies, there are few recent studies that investigated the impact of fasting in the month of Ramadan (this fasting lasts about 15 hours during the daytime) during the summer season, where temperatures reach a mean 45 ºC (SD 3 ºC) during the daytime and about 35 ºC (SD 5 ºC) during the night. Thus, the main aim of this study was to investigate the dynamics of blood lipids before, during, and after diurnal fasting in inactive men in hot environments.

## Methods

### Participants

This quasi-experimental pre-post design study was conducted at 4 different time points (before Ramadan, at week 2 and week 3 during Ramadan, and after Ramadan). A total of 44 inactive male students (mean age 27.6, SD 5.8 years) were recruited to participate in this study. Inclusion criteria were being male, an adult (aged 18-39 years), and inactive, that is, performing less than the recommended level of physical activity (PA) for adults (150 min of moderate PA per week, 75 min of vigorous PA per week, or an equivalent amount of moderate-to-vigorous PA per week). Smokers or individuals taking medicines that may have affected the results were excluded from participating in this study. The sample size was calculated based on the change in LDL concentration reported in a previous study [27] with the 95% CI and 80% power for the test. The estimated sample size was 38 participants. The number of recruited participants was increased by about 10% to prevent the results being affected by withdrawals or incomplete data; thus, 44 apparently healthy inactive male students were recruited to participate in this study.

### Ethical Considerations

A consent form was read and signed by each participant before taking part in this study. Ethical approval was obtained from the Deanship of Graduate Studies and Scientific Research at King Saud University (CSSPA-22-07-35). The participants did not receive any compensation. All data were saved anonymously in a safe and confidential format by the principal investigator.

### Procedures

Each participant was instructed by the principle investigator to visit the laboratory on the same day of the week on all 4 occasions. The measured variables were obtained at each visit. The first visit was 2 to 3 days before the month of Ramadan, and the second and third visits were in the second week and third week of Ramadan. The fourth visit took place 6 weeks after the month of Ramadan. The procedures of this study were performed consistently as described in the following sections.

### Physical Characteristics

Body mass (BM), height, and waist circumference (WC) were measured. BMI was calculated by dividing body weight in kilograms by height in meters squared ($kg/m^2$). Fat-free mass (FFM), body fat (BF) percentage, and total body water (TBW) percentage were evaluated using bioelectrical impedance analysis (BC-418; Tanita) [28].

### PA and Diet Assessment

PA and diet behavior were evaluated using a valid and reliable self-reported questionnaire for the assessment of PA and diet behavior in youth aged 15 to 25 years [29]. Each participant was instructed on how to fill out the PA and diet questionnaire. The questionnaire provided continuous variables (time in minutes). The average time to complete the questionnaire was 10 minutes. The diet behavior section assessed the type and frequency of the consumed food.

### Physiological and Lipid Profile

Resting heart rate (HR) and blood pressure (BP) were recorded following 10 minutes of being seated on a chair. All participants were instructed to fast at least 10 hours before attending the laboratory for blood collection. A venous blood sample of 8 ml was drawn by a phlebotomist to evaluate blood lipids, including TG, total cholesterol (TC), HDL, and LDL. Blood samples were analyzed in a specialized medical laboratory at the university.

### Statistical Analysis

Data from this study were analyzed using SPSS (version 24.0; IBM Corp). Results were illustrated as means with SDs. A 1-way repeated measures ANOVA was used to compare the measured variables before, during, and after the month of Ramadan. The Tukey post hoc test was used when a significant difference was found. The $\chi^2$ test was used to determine the statistical significance of associations between food consumed per week for participants across all trials. Results were considered statistically significant at $P<.05$.

## Results

The main aim of this study was to investigate the dynamics of blood lipids before, during, and after diurnal fasting in inactive men. Table 1 shows the physical characteristics of the participants before, during, and after Ramadan. This study did not observe significant changes in any of the body composition variables, including BM, BMI, WC, BF mass, BF percentage, FFM, FFM percentage, and TBW between all measured occasions.

**Table 1.** Measurements of body composition at the 4 time points (N=44).

| Body composition measurements | Before Ramadan, mean (SD) | Ramadan week 2, mean (SD) | Ramadan week 3, mean (SD) | After Ramadan, mean (SD) | *F* test (*df*) | *P* value |
|---|---|---|---|---|---|---|
| Body mass (kg) | 70.0 (12.6) | 69.6 (12.4) | 69.6 (12.8) | 69.9 (12.9) | 0.03 (43) | .99 |
| BMI (kg/m$^2$) | 24.5 (4.0) | 24.3 (4.0) | 24.3 (3.8) | 24.3 (4.0) | 0.04 (43) | .99 |
| Waist circumference (cm) | 82.9 (10.9) | 82.3 (10.6) | 81.8 (10.5) | 81.8 (11.0) | 0. 10 (43) | .96 |
| Body fat mass (kg) | 14.7 (6.6) | 14.0 (6.9) | 13.9 (6.8) | 14.5 (6.9) | 0.13 (43) | .94 |
| Body fat percentage (%) | 20.2 (6.2) | 19.5 (6.7) | 19.4 (6.6) | 19.7 (6.5) | 0.22 (43) | .88 |
| Fat-free mass (kg) | 55.3 (8.6) | 55.6 (6.9) | 55.5 (7.5) | 55.8 (7.3) | 0.04 (43) | .99 |
| Fat-free mass percentage (%) | 79.8 (6.2) | 80.5 (9.0) | 80.6 (6.7) | 80.3 (6.5) | 0.13 (43) | .94 |
| Total body water (liters) | 40.1 (6.0) | 41.6 (6.8) | 40.5 (5.8) | 40.6 (5.4) | 0.49 (43) | .69 |
| Total body water percentage (%) | 56.9 (7.2) | 59.0 (5.0) | 59.3 (5.3) | 58.9 (5.0) | 1.56 (43) | .20 |

### Physical Activity

Based on the PA guidelines (moderate-to-vigorous PA ≥150 min/week for ≥5 days/week or ≥70 min/week for ≥3 days/week) [30], participants were physically inactive. Moreover, there was no significant change in PA level across the 4 time points (99.8, SD 210.9; 125.7, SD 260.4; 111.0, SD 275.4; and 56.7, SD 151.0 minutes before Ramadan, in the second and third weeks of Ramadan, and after Ramadan, respectively; Table 2). The average sedentary time of the participants was less than 6 hours (ranging from 5.2, SD 1.9 to 5.8, SD 2.2 h/day) across all 4 time points.

**Table 2.** Physical activity levels across all 4 time points (before Ramadan, in the second and third weeks of Ramadan, and after Ramadan; N=44).

| Physical activity and sleeping time | Before Ramadan, mean (SD) | Ramadan week 2, mean (SD) | Ramadan week 3, mean (SD) | After Ramadan, mean (SD) | *F* test (*df*) | *P* value |
|---|---|---|---|---|---|---|
| Moderate activity (min/wk) | 50.9 (121.7) | 64.4 (137.6) | 50.2 (154.51) | 34.6 (89.5) | 0.40 (43) | .75 |
| Vigorous activity (min/wk) | 48.9 (108.4) | 61.3 (143.9) | 60.8 (162.4) | 22.2 (76.0) | 0.92 (43) | .43 |
| Moderate + vigorous (min/wk) | 99.8 (210.9) | 125.7 (260.4) | 111.0 (275.4) | 56.7 (151.0) | 0.74 (43) | .53 |
| Sedentary behavior (h/day) | 5.8 (2.2) | 5.2 (1. 9) | 5.7 (3.0) | 5.5 (2.6) | 0.56 (43) | .64 |
| Sleeping duration (h/day) | 6.9 (1.0) | 6.6 (1.1) | 6.7 (1.0) | 6.3 (1.1) | 2.58 (43) | .06 |

### Diet Behavior

Diet behavior was reported by evaluating types and frequencies of consumed food. Diet behavior is one of the essential factors that could have impacted this study's outcomes. Thus, participants' diet behavior was evaluated across all 4 time points. Table 3 demonstrates the consumed food per week at all 4 time points. The frequency of the consumed food was classified into 3 categories: no more than twice per week, 3 to 5 times per week, and ≥6 times per week. There were no significant changes in the frequency of consumed food per week across the 4 time points.

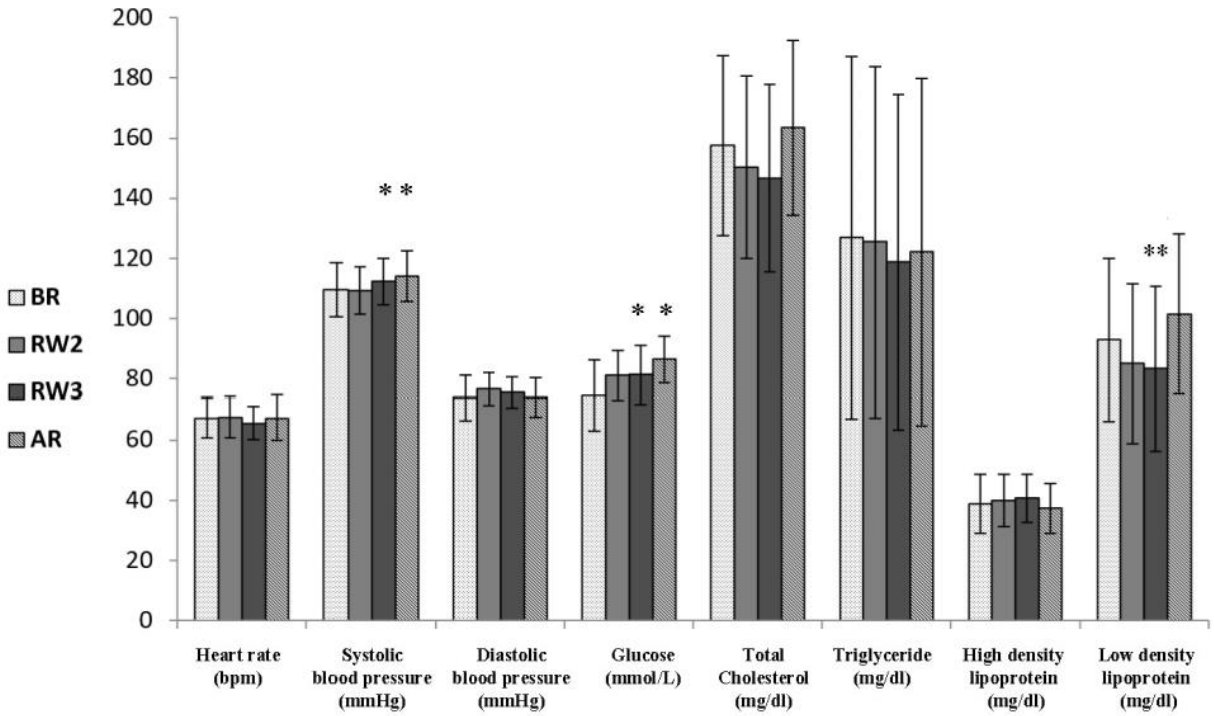
**Table 3.** Percentage of consumed food per week for participants across all 4 time points (before Ramadan, in the second and third weeks of Ramadan, and after Ramadan; N=44).

| Food types | Frequency of consumed food per week (participants), n (%) | | | Chi-square (*df*) | *P* value |
|---|---|---|---|---|---|
| | <2 times/w | <3-5 times/w | >6 times/w | | |
| Vegetables | 12 (28) | 17 (40) | 14 (32) | 3.97 (43) | .68 |
| Fruit | 9 (21) | 21 (48) | 14 (32) | 9.11 (43) | .17 |
| Milk and dairy products | 13 (30) | 18 (40) | 13 (30) | 5.62 (43) | .47 |
| Soda and soft drinks | 19 (44) | 15 (34) | 10 (22) | 11.37 (43) | .08 |
| Energy drinks | 39 (89) | 4 (10) | 2 (1) | 3.90 (43) | .27 |
| Fast food | 19 (43) | 18 (43) | 6 (14) | 6.20 (43) | .40 |
| Chips and french fries | 20 (46) | 17 (40) | 6 (14) | 3.70 (43) | .72 |
| Biscuits and cake | 24 (55) | 11 (26) | 9 (20) | 11.97 (43) | .22 |
| Chocolate and sweets | 16 (36) | 13 (31) | 14 (33) | 9.30 (43) | .16 |

## Physiological and Blood Parameters

Physiological and blood parameters, including resting HR, BP, and blood lipids, were analyzed. The results of this study showed a significant effect of diurnal fasting on both FBG and LDL (Figure 1). Furthermore, a significant elevation was recorded in systolic blood pressure (SBP) 1 month after Ramadan compared to before Ramadan (109 mm Hg before Ramadan vs 114 mm Hg after Ramadan). Similarly, FBG was slightly elevated during week 3 of Ramadan and after Ramadan compared to before Ramadan (74.60 mmol/L before Ramadan, 81.52 mmol/L at week 3, and 86.51 mmol/L after Ramadan). However, the results showed an obvious reduction in LDL during week 3 of Ramadan compared to before and after Ramadan month (83.49 mg/dl at week 3 vs 93.11 mg/dl before Ramadan [*P*=.02] and 101.59 mg/dl after Ramadan [*P*=.007]).

**Figure 1.** The effect of Ramadan fasting on resting heart rate (HR), systolic blood pressure, diastolic blood pressure, glucose, total cholesterol, triglyceride, high density lipoprotein (HDL) and low-density lipoprotein (LDL), (n=44).



## Discussion

### Principal Findings

The impact of diurnal fasting during the month of Ramadan on blood lipids in inactive men was limited but positive and very obvious for LDL. Although there were no significant changes in PA or SB across all trials ($P$=.53 for moderate-to-vigorous PA; $P$=.64 for SB), some blood lipids, such as LDL, may be affected positively during fasting in the month of Ramadan. The novelty of this study was controlling for the main related factors that may affect blood lipid variables. In contrast, this study did not find significant changes in the participants' diet behavior across the investigated time points (all $P$>.05). In general, the findings of this study may help inactive individuals control blood lipids, especially LDL, while fasting diurnally.

### Body Composition

All body composition variables investigated in this study did not change significantly across any of the 4 time points. The novelty of this study was examining the impact of Ramadan fasting in a hot and dry environment (≥45 ºC) on blood lipids. In this study, body mass, BMI, and WC decreased gradually, especially during the third week of Ramadan, compared to before Ramadan. However, the recorded reduction was not significant ($P$=.88 for BF percentage, $P$=.94 for FFM percentage, $P$=.20 for BW percentage, $P$=.99 for both BM and BMI, and $P$=.96 for WC). While some previous studies have reported significant decreases in body composition parameters at the end of Ramadan compared to measurements taken before Ramadan [21], numerous previous studies have concurred with the findings of this study [7,25,26,31-33]. In fact, the results of this study are reasonable, as most of the participants were within normal weight. However, some studies have reported a reduction in BW due to a decrease in skeletal muscle mass and FFM [34]. Furthermore, the significant reduction in body composition in the previous literature may be linked to the obesity status of the participants [35]. The majority of the participants in this study were within normal body composition measurements. Therefore, the impact of Ramadan fasting may not influence body composition in individuals with normal body composition [36]. Focusing on apparently healthy men could be one of the limitations of this study. Further studies in different ages and genders are encouraged to understand the impact of diurnal fasting on body composition parameters.

### Physiological and Blood Parameters

The results for both SBP and FBG were elevated significantly during the month of Ramadan ($P$=.02 and $P$=.01, respectively), especially during the end of Ramadan (week 3) compared to levels before Ramadan. These findings concur with previous studies [22,37]. However, the elevated values of SBP and FBG were within the healthy range (SBP: 109 mm Hg before Ramadan vs 114 mm Hg after Ramadan; FBG: 74.60 mmol/L before Ramadan vs 81.52 mmol/L at week 3 and 86.51 mmol/L after Ramadan). One novel aspect of this study was the impact of diurnal fasting on LDL, as there were no significant changes in the main factors that may impact LDL concentration, such as body composition, PA, and diet behavior. Thus, diurnal fasting itself could be an independent factor that positively improved LDL. Our findings concur with the results of a previous study that investigated the impact of Ramadan fasting on college students [23,36,38]. However, it seems that the observed positive, significant change in LDL concentration as a result of diurnal fasting during the month of Ramadan was temporary [25,38]. The novelty of this study is its examination of the impact of diurnal fasting on blood lipids in inactive men with a stable lifestyle. Blood biomarkers, including blood lipid concentration, may be positively improved by diurnal fasting. However, this study discovered that the changes in blood lipid variables were limited in healthy but inactive young men. The major changes in blood parameters observed in previous studies may refer to the nature of participants' lifestyles [22,24,39]. Further investigations are encouraged to clarify the impact of diurnal fasting on blood lipids in people with special conditions, such as older age and obesity.

This study has some limitations. First, it was conducted with apparently healthy young adult men. Second, physical activity was not assessed by an objective measurement such as an accelerometer. Third, some other confounders were not evaluated, such as the total energy expenditure and basal metabolic rate of the participants. Moreover, the sample size of the study may not have been enough to detect differences between the time points.

### Conclusion

Diurnal intermittent fasting seems to be an independent factor that could improve blood lipids, including LDL, even in healthy but inactive adults. Moreover, diurnal fasting has been found to decrease LDL even in inactive adults with normal LDL levels. However, the results of this study concur with previous studies in terms of the temporary effect of diurnal fasting during the month of Ramadan on blood lipids. More research is encouraged to examine the impact of Ramadan fasting in people with special health conditions, such as obesity or older age.

### Acknowledgments

The authors would like to extend their gratitude to the Researchers Supporting Project (RSP2024R263), King Saud University, Riyadh, Saudi Arabia.

### Data Availability

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

## Authors' Contributions

**Conceptualization –** KA and NA-B
**Methodology –** KA and NA-B
**Validation –** KA, NA-B, AN, and GA
**Formal analysis –** KA, NA-B, and AA
**Investigation –** KA, NA-B, AA, YA, and GA
**Writing – original draft preparation –** KA, NA-B, and GA
**Writing – review & editing –** KA, NA-B, AN, AA, YA, and GA
**Supervision –** KA and GA.

## Conflicts of Interest

None declared.

## References


1. Abdelrahim D, Faris ME, Hassanein M, Shakir AZ, Yusuf AM, Almeneessier AS, et al. Impact of Ramadan diurnal intermittent fasting on hypoglycemic events in patients with type 2 diabetes: a systematic review of randomized controlled trials and observational studies. Front Endocrinol (Lausanne) 2021;12:624423 [FREE Full text] [doi: 10.3389/fendo.2021.624423] [Medline: 33763029]
2. Grim BJ, Hsu B. Estimating the global Muslim population: size and distribution of the world's Muslim population. Interdiscip J Res Relig 2011;7(2):1-19 [FREE Full text]
3. Baanter NS, Elsayed AAA, Aljerian K, Aldahmash WM, Al-Qahtani WS, Alwasel SH. Changes in maternal lifestyle during Ramadan altered placental development. Adv Reprod Sci 2020;8(2):126-142. [doi: 10.4236/ARSCI.2020.82011]
4. Ünal K, Topcuoglu C, Yucel C, Turhan T. Effects of Ramadan lifestyle on lipid profile and oxidative stress markers in adult males. J Clin Anal Med 2019;10(1):5-9. [doi: 10.4328/JCAM.5943]
5. Aljaloud K. The effect of Ramadan fasting on metabolic syndrome (MetS). In: Szablewski J, editor. Blood Glucose Levels. London, UK: IntechOpen; 2020.
6. Ranasinghe P, Mathangasinghe Y, Jayawardena R, Hills AP, Misra A. Prevalence and trends of metabolic syndrome among adults in the Asia-Pacific region: a systematic review. BMC Public Health 2017 Jan 21;17(1):101 [FREE Full text] [doi: 10.1186/s12889-017-4041-1] [Medline: 28109251]
7. Sadiya A, Ahmed, Siddieg, Joy, Carlsson. Effect of Ramadan fasting on metabolic markers, body composition, and dietary intake in Emiratis of Ajman (UAE) with metabolic syndrome. Diabetes Metab Syndr Obes 2011 Dec;4:409. [doi: 10.2147/dmso.s24221]
8. Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. JAMA 2002 Jan 16;287(3):356-359. [doi: 10.1001/jama.287.3.356] [Medline: 11790215]
9. King R, Rotter J, Motulsky A. Approach to genetic bases of common diseases. In: The Genetic Basis of Common Diseases. Oxford, UK: Oxford University Press; 2002:3-17.
10. Aktas MF, Mähler A, Hamm M, Perger G, Simon F, Westenhöfer J, et al. Lifestyle interventions in Muslim patients with metabolic syndrome-a feasibility study. Eur J Clin Nutr 2019 May;73(5):805-808. [doi: 10.1038/s41430-018-0371-z] [Medline: 30538299]
11. Faris, Jahrami HA, Alsibai J, Obaideen AA. Impact of Ramadan diurnal intermittent fasting on the metabolic syndrome components in healthy, non-athletic Muslim people aged over 15 years: a systematic review and meta-analysis. Br J Nutr 2020 Jan 14;123(1):1-22. [doi: 10.1017/S000711451900254X] [Medline: 31581955]
12. Jakicic J, Donnelly JE, Jawad AF, Jacobsen DJ, Gunderson SC, Pascale R. Association between blood lipids and different measures of body fat distribution: effects of BMI and age. Int J Obes Relat Metab Disord 1993 Mar;17(3):131-137. [Medline: 8385072]
13. Cho ER, Shin A, Kim J, Jee SH, Sung J. Leisure-time physical activity is associated with a reduced risk for metabolic syndrome. Ann Epidemiol 2009 Nov;19(11):784-792. [doi: 10.1016/j.annepidem.2009.06.010] [Medline: 19825459]
14. Rokling-Andersen MH, Reseland JE, Veierød MB, Anderssen SA, Jacobs DR, Urdal P, et al. Effects of long-term exercise and diet intervention on plasma adipokine concentrations. Am J Clin Nutr 2007 Nov;86(5):1293-1301 [FREE Full text] [doi: 10.1093/ajcn/86.5.1293] [Medline: 17991638]
15. Kondo T, Kobayashi I, Murakami M. Effect of exercise on circulating adipokine levels in obese young women. Endocr J 2006 Apr;53(2):189-195 [FREE Full text] [doi: 10.1507/endocrj.53.189] [Medline: 16618976]
16. Al-Hazzaa H. Physical inactivity in Saudi Arabia revisited: A systematic review of inactivity prevalence and perceived barriers to active living. Int J Health Sci (Qassim) 2018;12(6):50-64 [FREE Full text] [Medline: 30534044]
17. Hamer M, O'Donovan G, Murphy M. Physical inactivity and the economic and health burdens due to cardiovascular disease: exercise as medicine. Adv Exp Med Biol 2017;999:3-18. [doi: 10.1007/978-981-10-4307-9_1] [Medline: 29022254]

18. Bibliometric overview of Ramadan fasting studies during 2010-2021. Diabetes Metab Syndr 2022 Jul;16(7):102531. [doi: 10.1016/j.dsx.2022.102531] [Medline: 35717895]
19. Bandarian F, Namazi N, Atlasi R, Nasli-Esfahani E, Larijani B. Research gaps in Ramadan fasting studies in health and disease. Diabetes Metab Syndr 2021;15(3):831-835. [doi: 10.1016/j.dsx.2021.04.009] [Medline: 33866118]
20. Darzabi T, Hejazi K, Shahrokhi H. The effects of Ramadan fasting on metabolic syndrome and body composition in male students. J Client-Centered Nurs Care 2019 Nov 01;5(4):239-246. [doi: 10.32598/JCCNC.5.4.311.1]
21. Nachvak SM, Pasdar Y, Pirsaheb S, Darbandi M, Niazi P, Mostafai R, et al. Effects of Ramadan on food intake, glucose homeostasis, lipid profiles and body composition composition. Eur J Clin Nutr 2019 Apr;73(4):594-600. [doi: 10.1038/s41430-018-0189-8] [Medline: 29884874]
22. Ongsara S, Boonpol S, Prompalad N, Jeenduang N. The effect of Ramadan fasting on biochemical parameters in healthy Thai subjects. J Clin Diagn Res 2017 Sep;11(9):BC14-BC18. [doi: 10.7860/jcdr/2017/27294.10634]
23. Saleh Mansi KM. Study the effects of Ramadan fasting on the serum glucose and lipid profile among healthy Jordanian students. Am J Appl Sci 2007 Aug 1;4(8):565-569. [doi: 10.3844/AJASSP.2007.565.569]
24. Waterhouse J, Alabed H, Edwards B, Reilly T. Changes in sleep, mood and subjective and objective responses to physical performance during the daytime in Ramadan. Biol Rhythm Res 2009 Aug 27;40(5):367-383. [doi: 10.1080/09291010902863438]
25. Leiper JB, Molla AM, Molla AM. Effects on health of fluid restriction during fasting in Ramadan. Eur J Clin Nutr 2003 Dec;57 Suppl 2:S30-S38. [doi: 10.1038/sj.ejcn.1601899] [Medline: 14681711]
26. Ramadan J. Does fasting during Ramadan alter body composition, blood constituents and physical performance? Med Princ Pract 2002;11(Suppl 2):41-46 [FREE Full text] [doi: 10.1159/000066413] [Medline: 12444309]
27. Ziaee V, Razaei M, Ahmadinejad Z, Shaikh H, Yousefi R, Yarmohammadi L, et al. The changes of metabolic profile and weight during Ramadan fasting. Singapore Med J 2006 May;47(5):409-414 [FREE Full text] [Medline: 16645692]
28. Jaffrin M. Body composition determination by bioimpedance: an update. Curr Opin Clin Nutr Metab Care 2009 Sep;12(5):482-486. [doi: 10.1097/MCO.0b013e32832da22c] [Medline: 19494768]
29. Al-Hazzaa H, Al-Ahmadi M. A self-reported questionnaire for the assessment of physical activity in youth 15-25 years: development, reliability and construct validity. Arab J Food Nutr 2003;4:279-291 [FREE Full text]
30. Elgaddal N, Kramarow EA, Reuben C. Physical activity among adults aged 18 and over: United States, 2020. NCHS Data Brief 2022 Aug(443):1-8. [doi: 10.15620/cdc:120046] [Medline: 36043905]
31. Al-Hourani H, Atoum MF. Body composition, nutrient intake and physical activity patterns in young women during Ramadan. Singapore Med J 2007 Oct;48(10):906-910 [FREE Full text] [Medline: 17909674]
32. Sadeghirad B, Motaghipisheh S, Kolahdooz F, Zahedi MJ, Haghdoost AA. Islamic fasting and weight loss: a systematic review and meta-analysis. Public Health Nutr 2014 Feb;17(2):396-406 [FREE Full text] [doi: 10.1017/S1368980012005046] [Medline: 23182306]
33. Asl NS. The effects of Ramadan fasting on endurance running performance in male athletes. Int J Sport Stud 2011 2011;1:18-22.
34. Nugraha B, Ghashang SK, Hamdan I, Gutenbrunner C. Effect of Ramadan fasting on fatigue, mood, sleepiness, and health-related quality of life of healthy young men in summer time in Germany: A prospective controlled study. Appetite 2017 Apr 01;111:38-45. [doi: 10.1016/j.appet.2016.12.030] [Medline: 28027907]
35. Sezen Y, Altiparmak IH, Erkus ME, Kocarslan A, Kaya Z, Gunebakmaz O, et al. Effects of Ramadan fasting on body composition and arterial stiffness. J Pak Med Assoc 2016 Dec;66(12):1522-1527. [Medline: 27924959]
36. Harder-Lauridsen NM, Rosenberg A, Benatti FB, Damm JA, Thomsen C, Mortensen EL, et al. Ramadan model of intermittent fasting for 28 d had no major effect on body composition, glucose metabolism, or cognitive functions in healthy lean men. Nutrition 2017 May;37:92-103 [FREE Full text] [doi: 10.1016/j.nut.2016.12.015] [Medline: 28359370]
37. Al-Numair K. Body weight and some biochemical changes associated with Ramadan fasting in healthy Saudi men. Egypt J Agric Res 2006 Jan 01;57(1):67-78. [doi: 10.21608/ejarc.2006.224825]
38. Qujeq D, Bijani K, Kalavi K, Mohiti J, Aliakbarpour H. Effects of Ramadan fasting on serum low-density and high-density lipoprotein-cholesterol concentrations. Ann Saudi Med 2002;22(5-6):297-299 [FREE Full text] [doi: 10.5144/0256-4947.2002.297] [Medline: 17146246]
39. Bener A, A Al-Hamaq AOA, Öztürk M, Çatan F, Haris P, Rajput K, et al. Effect of Ramadan fasting on glycemic control and other essential variables in diabetic patients. Ann Afr Med 2018;17(4):196-202 [FREE Full text] [doi: 10.4103/aam.aam_63_17] [Medline: 30588933]


## Abbreviations

**BF:** body fat
**BM:** body mass
**BP:** blood pressure
**CVD:** cardiovascular disease
**FBG:** fasting blood glucose

**FFM:** fat-free mass
**HDL:** high-density lipoprotein
**HR:** heart rate
**LDL:** low-density lipoprotein
**PA:** physical activity
**SB:** sedentary behavior
**SBP:** systolic blood pressure
**TBW:** total body water
**TC:** total cholesterol
**TG:** triglyceride
**WC:** waist circumference

*Edited by T de Azevedo Cardoso; submitted 09.01.24; peer-reviewed by Y Cui, AMA Ismail, SM Farid Hasan; comments to author 28.05.24; revised version received 12.06.24; accepted 29.08.24; published 17.10.24.*

*Please cite as:*
*Aljaloud K, Al-Barha N, Noman A, Aldayel A, Alsharif Y, Alshuwaier G*
*Dynamics of Blood Lipids Before, During, and After Diurnal Fasting in Inactive Men: Quasi-Experimental Study*
*Interact J Med Res 2024;13:e56207*
*URL: https://www.i-jmr.org/2024/1/e56207*
*doi:10.2196/56207*
*PMID:*

Review

# Benefits and Risks of AI in Health Care: Narrative Review

Margaret Chustecki[1], MBA, MD

Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States

**Corresponding Author:**
Margaret Chustecki, MBA, MD
Department of Internal Medicine
Yale School of Medicine
1952 Whitney Ave
3rd Floor
New Haven, CT, 06510
United States
Phone: 1 2038091700
Email: margaret.chustecki@imgnh.com

## Abstract

**Background:** The integration of artificial intelligence (AI) into health care has the potential to transform the industry, but it also raises ethical, regulatory, and safety concerns. This review paper provides an in-depth examination of the benefits and risks associated with AI in health care, with a focus on issues like biases, transparency, data privacy, and safety.

**Objective:** This study aims to evaluate the advantages and drawbacks of incorporating AI in health care. This assessment centers on the potential biases in AI algorithms, transparency challenges, data privacy issues, and safety risks in health care settings.

**Methods:** Studies included in this review were selected based on their relevance to AI applications in health care, focusing on ethical, regulatory, and safety considerations. Inclusion criteria encompassed peer-reviewed articles, reviews, and relevant research papers published in English. Exclusion criteria included non–peer-reviewed articles, editorials, and studies not directly related to AI in health care. A comprehensive literature search was conducted across 8 databases: OVID MEDLINE, OVID Embase, OVID PsycINFO, EBSCO CINAHL Plus with Full Text, ProQuest Sociological Abstracts, ProQuest Philosopher's Index, ProQuest Advanced Technologies & Aerospace, and Wiley Cochrane Library. The search was last updated on June 23, 2023. Results were synthesized using qualitative methods to identify key themes and findings related to the benefits and risks of AI in health care.

**Results:** The literature search yielded 8796 articles. After removing duplicates and applying the inclusion and exclusion criteria, 44 studies were included in the qualitative synthesis. This review highlights the significant promise that AI holds in health care, such as enhancing health care delivery by providing more accurate diagnoses, personalized treatment plans, and efficient resource allocation. However, persistent concerns remain, including biases ingrained in AI algorithms, a lack of transparency in decision-making, potential compromises of patient data privacy, and safety risks associated with AI implementation in clinical settings.

**Conclusions:** In conclusion, while AI presents the opportunity for a health care revolution, it is imperative to address the ethical, regulatory, and safety challenges linked to its integration. Proactive measures are required to ensure that AI technologies are developed and deployed responsibly, striking a balance between innovation and the safeguarding of patient well-being.

***(Interact J Med Res 2024;13:e53616)*** doi:10.2196/53616

**KEYWORDS**

artificial intelligence; safety risks; biases; AI; benefit; risk; health care; safety; ethics; transparency; data privacy; accuracy

## Introduction

Artificial intelligence (AI) has rapidly proliferated across various sectors in recent years, with the health care industry emerging as a primary arena for its transformative potential. This technological advancement holds promise for revolutionizing patient care and administrative operations by leveraging vast longitudinal patient data [1]. AI encompasses a spectrum of technologies, including machine learning (ML), natural language processing (NLP), rule-based expert systems (RBES), physical robots, and robotic process automation, each offering unique capabilities from predictive modeling and disease detection to enhancing surgical precision and automating administrative

tasks [2-7]. The integration of AI into health care promises heightened diagnostic accuracy, informed decision-making, and optimized treatment planning, thereby potentially reducing medical errors and improving patient outcomes [1].

However, alongside these promising developments, AI adoption in health care is accompanied by significant ethical and regulatory challenges that require careful consideration [8]. Concerns range from safeguarding patient data privacy to addressing algorithmic biases that may perpetuate disparities in health care outcomes [9,10]. The regulatory landscape is evolving to keep pace with technological advancements, aiming to establish robust governance frameworks that ensure the responsible use of AI in health care settings. Furthermore, the advent of pretrained large language models, exemplified by models like BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer), and their variants, has further expanded the capabilities of AI in health care [11-14]. These models leverage vast amounts of text data to learn rich representations of language, enabling tasks ranging from clinical documentation improvement to automated summarization of medical literature [15,16].

Against this backdrop, this study presents a narrative review aimed at comprehensively exploring the multifaceted role of AI in health care. By synthesizing existing literature, this research aims to provide insights into the diverse applications of AI, its associated benefits, and the ethical and regulatory considerations that underpin its integration into clinical practice [9,10,17]. This review aims to facilitate informed decision-making among health care professionals, policy makers, and researchers, fostering a balanced approach that maximizes the benefits of AI while mitigating potential risks within the health care landscape.

This review seeks to contribute to ongoing discussions on AI ethics, governance, and effective deployment strategies, thereby guiding the responsible and impactful adoption of AI technologies in health care. By examining current trends, challenges, and future directions, this review aims to lay the groundwork for advancing AI's role in enhancing health care delivery, improving patient outcomes, and supporting health care systems globally.

## Methods

### Overview

This narrative review aims to assess the benefits and risks associated with the integration of AI into health care, with a primary focus on potential biases, transparency issues, data privacy concerns, and safety risks. A literature review was conducted to explore the current landscape of AI applications in health care and to identify relevant ethical, regulatory, and safety considerations.

### Eligibility Criteria

Specific inclusion and exclusion criteria were established to guide the selection of studies for this narrative review. Studies were included if they were relevant to the 3 core concepts of AI, ethics, and health and were written in the English language. Articles were excluded if they did not explicitly address each of the core concepts of AI, ethics, and health or if they were not written in English. In addition, studies focusing solely on ethics and big data without explicit mention of AI methods or applications were excluded. Non–peer-reviewed academic literature, such as letters and non–peer-reviewed conference proceedings, as well as books and book chapters, were also excluded as they were deemed irrelevant to this review. No restrictions were applied regarding the publication date or study design to ensure a broad overview of the topic.

### Information Sources

The literature search used 8 electronic databases: OVID MEDLINE (1946-present), OVID Embase (1947-present), OVID PsycINFO (1806-present), EBSCO CINAHL Plus with Full Text (1937-present), ProQuest Sociological Abstracts (1952-present), ProQuest Philosopher's Index (1940-present), ProQuest Advanced Technologies & Aerospace (1962-present), and Wiley Cochrane Library. Search strategies were tailored to each database (Multimedia Appendix 1), using controlled vocabulary, Medical Subject Headings (MeSH) terms, EMTREE terms, American Psychological Association's Thesaurus of Psychological Index Terms, CINAHL headings, Sociological Thesaurus, Philosopher's Index subject headings, and Advanced Technologies & Aerospace subject headings. The searches were limited to English language–only articles, and filters excluding animal studies were applied to specific databases. In addition, a filter for health or medicine-related studies was applied to the Advanced Technologies & Aerospace database.

The final searches of the peer-reviewed literature were completed on June 23, 2023. Gray literature was not searched in this narrative review.

### Selection and Sources of Evidence

All identified records from the academic literature searches were imported into the reference management software EndNote (Clarivate). After removing duplicate records, screening was conducted in 2 steps: initial title and abstract screening followed by full-text review. Full-text reviews were conducted to ensure that the selected studies provided substantial insights for the narrative synthesis.

### Data Charting Process

Data charting forms were developed and refined based on the narrative review research question. The forms included fields for recording data such as the objective of each paper, institutional affiliations of authors, publication year, country of the first and corresponding authors, conflict of interest disclosures, health context of interest, AI applications or technologies discussed, ethical concepts, issues or implications raised, reference to global health, and recommendations for future research, policy, or practice. Data were recorded directly into the data charting form with corresponding page numbers to ensure accuracy.

### Synthesis of Results

Data analysis included thematic components. Thematic analysis was conducted inductively, generating open descriptive codes from a sample of records. Codes were applied to relevant data points across all records, with new codes added as needed. These

codes were then organized into themes, allowing for the identification of commonalities and gaps in the literature. Results are presented in a narrative format.

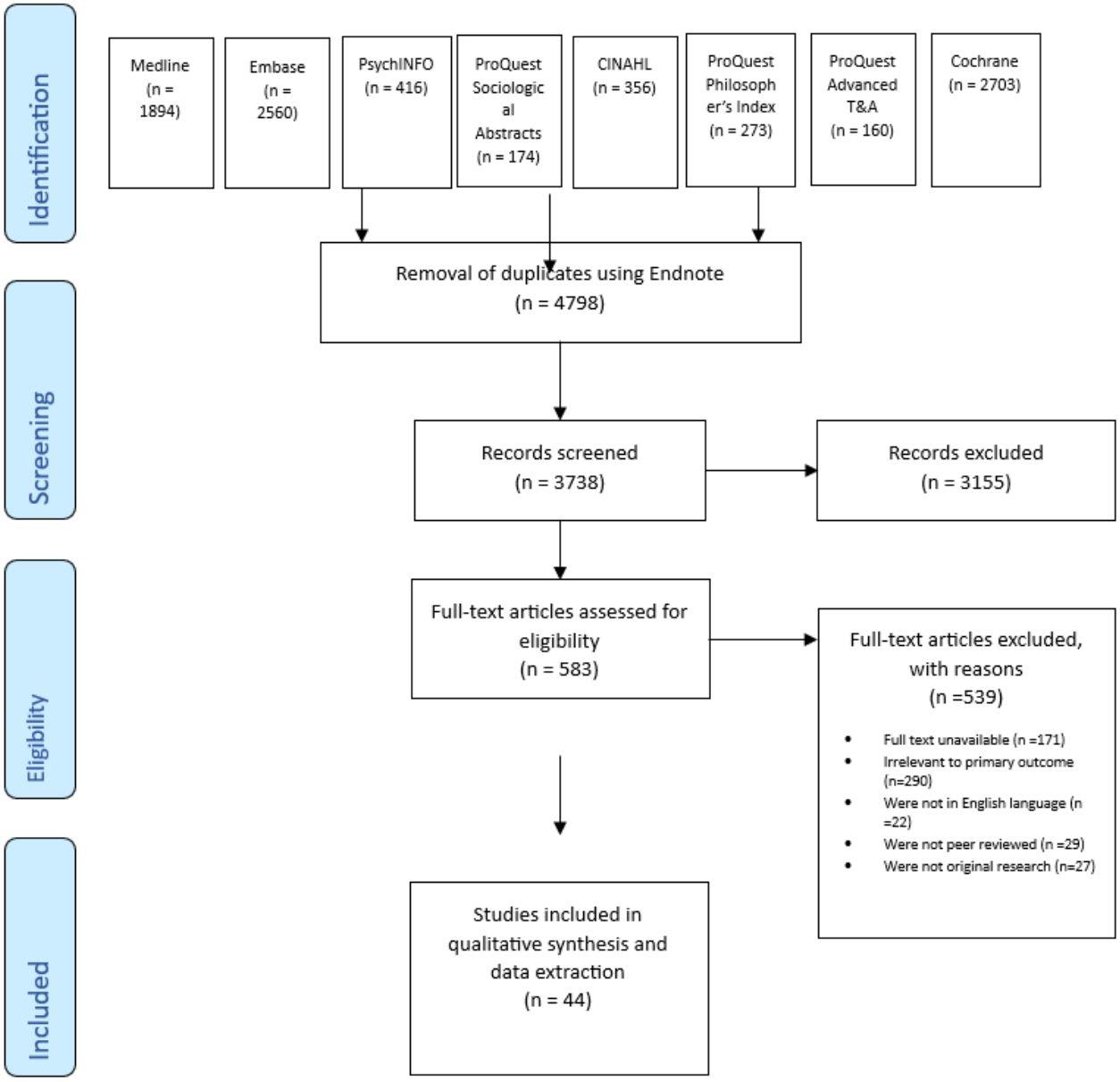
## Results

### Overview

Within the realm of integrating AI into health care, this narrative review has revealed a broad range of insights that span a spectrum of possibilities and challenges. This section categorizes the findings into 2 overarching categories: “Benefits” and “Risks.” Each category encapsulates a tapestry of themes that emerged from an exploration of academic literature. As these themes are explored, the multifaceted landscape of AI’s influence on health care is illuminated. The “Benefits” section unveils the potential for AI to revolutionize health care delivery, ushering in more accurate diagnoses, personalized treatment regimens, and streamlined resource allocation. Conversely, the “Risks” section delves into the intricate ethical, privacy, and safety concerns that accompany the integration of AI into clinical settings. Through a comprehensive examination of these themes, this review provides a nuanced perspective on the implications and imperatives in harnessing AI’s potential for the betterment of health care.

The systematic literature review process, as illustrated in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1), outlines a thorough and rigorous methodology. Initial searches across multiple databases—MEDLINE, Embase, PsycINFO, ProQuest Sociological Abstracts, CINAHL, ProQuest Philosopher’s Index, ProQuest Advanced Technologies & Aerospace, and Cochrane Library—yielded a total of 8796 articles. After removing 4798 duplicates using Endnote, 3738 unique records were screened for relevance. Of these, 3155 articles were excluded based on title and abstract review for not meeting the inclusion criteria. The remaining 583 articles underwent full-text assessment for eligibility. Further evaluation led to the exclusion of 539 articles due to various reasons, such as unavailability of full text (n=171), irrelevance to the primary outcome (n=290), non-English language (n=22), not being peer-reviewed (n=29), and not being original research (n=27). Ultimately, 44 studies were included in the qualitative synthesis and data extraction. This meticulous selection process ensured that the final set of studies provided a robust and representative foundation for examining the integration of AI in health care.

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) study selection flow diagram outlining the literature review process when searching for articles on various databases.




## Benefits of AI in Health Care

Textbox 1 below describes the main benefits of implementing AI in health care. The benefits are explained in detail below.

**Textbox 1.** Benefits of artificial intelligence (AI) in health care.

**Medical benefits**

- Helps in prediction of various risks and diseases
- Helps in prevention and control of various diseases
- Leads to better data-driven decisions within the health care system
- Assists in improving surgery
- Supports mental health

**Economic and social benefits**

- Reduction in posttreatment expenditures
- Cost saving through early diagnosis
- Cost saving with enhanced clinical trials
- Patient empowerment
- Relieving medical practitioners' workload

## Medical Benefits

AI adoption in health care offers various medical, economic, and social benefits. This section discusses some of the key medical benefits of AI.

### *Prediction of Risks and Diseases*

AI leverages big data to predict diseases and assess risk exposure among patients. For example, Google collaborates with health delivery networks to develop prediction models that alert clinicians of high-risk illnesses like sepsis and heart failure [18]. ML models can also be used to forecast populations at risk of specific diseases or accidents [19,20]. In addition, AI algorithms, such as deep learning, aid in disease classification and enable more personalized care [21].

### *Prevention and Control of Diseases*

AI can play a significant role in the prevention and control of diseases. For instance, AI can enhance sexually transmitted infection (STI) prevention and control by improving surveillance and intervention. By analyzing publicly available social media data, AI can predict county-level syphilis prevalence, enabling faster and more efficient monitoring [22]. AI can also analyze trends in web data to reduce the stigma associated with STI prevention and care and identify and flag STI-related misinformation [22].

### *Data-Driven Decision Making*

AI enables better data-driven decisions within the health care system. In a digitalized health care environment, the quality of decision-making relies on the availability and accuracy of underlying data [23]. AI can assist in decision-making by offering real-time recommendations based on clinical guidelines or advancements, reducing the likelihood of medical mistakes [24]. For example, IBM Watson Health uses ML to provide clinical decision support and achieved a high level of agreement with physician recommendations [25].

### *Improvement in Surgery*

AI has made significant advancements in surgical procedures. Robotic surgery, such as in gynecologic, prostate, and oral and maxillofacial surgery, enhances surgical precision and predictability [7,26]. Telesurgical techniques driven by AI enable remote surgery and provide better supervision of surgeons [27]. AI-powered surgical mentorship allows skilled surgeons to offer real-time advice and guidance to other surgeons during procedures, improving surgical outcomes [28].

### *Mental Health Support.*

AI use in mental health treatments is growing as patients prefer simple and quick feedback [29]. According to Lovejoy et al [30], psychiatric professionals have historically relied on therapeutic discourse and patient narrative to assess mental health since language is the primary means to communicate our emotional and mental well-being. Recent advancements in AI have opened up new perspectives on the subject by enabling technology to infer emotional meaning from more data sources [30]. According to Habermann [31], with a unique combination of NLP and sentiment analysis, data scientists have developed algorithms to comprehend human emotion from the text. Le Glaz et al [32] mentions that in recent years, NLP models have been used to track mental self-disclosure on Twitter, forecast suicide risk online, and identify suicidal thoughts in clinical notes. These models are used in medicine to give complete details about a patient's emotional and psychological health [31].

## Economic and Social Benefits

In addition to the medical health benefits, using AI in health care has other economic and social advantages, as discussed below.

### *Reduction in Posttreatment Expenditures.*

AI-powered systems can analyze posttreatment result patterns and identify the most effective remedies based on patient profiles. This personalized approach to care can significantly reduce the expenses associated with posttreatment complications, which are a major cost driver in health care

systems worldwide [33]. By providing immediate diagnosis and appropriate interventions, AI can help minimize the financial burden of posttreatment complications and lead to substantial cost savings.

### *Cost Saving Through Early Diagnosis.*

AI has demonstrated superior accuracy and speed in analyzing medical images, such as mammograms, leading to the early detection of diseases like breast cancer. By enabling prompt diagnosis and action before issues escalate, AI can help reduce health care costs associated with late-stage diagnoses [28]. In addition, AI's ability to process and interpret various medical tests, such as computed tomography scans, with high accuracy reduces the likelihood of physician errors, contributing to cost savings.

### *Cost Saving with Enhanced Clinical Trials.*

AI-powered programs can simulate and evaluate numerous potential treatments to predict their effectiveness against various diseases, optimizing the drug development process in clinical trials [34]. By leveraging biomarker monitoring frameworks and analyzing large volumes of patient data, AI accelerates the evaluation of potential treatments, leading to significant cost savings in the development of life-saving medications.

### *Patient Empowerment.*

AI has the potential to empower individuals in managing their health. Wearable devices, such as smartwatches, can collect standard health data, which AI algorithms can analyze to provide personalized health recommendations and warnings for potential diseases [35]. Smartphone apps that use ML algorithms can help patients with chronic diseases better manage their conditions, leading to healthier populations and reduced health care expenses [36].

### *Relieving Medical Practitioners' Workload.*

AI technologies can alleviate the burden on health care workers by assisting with administrative tasks, data analysis, and image interpretation. AI can automate clerical responsibilities, analyze patient data more efficiently, and aid in diagnosing various medical conditions [37,38]. By reducing manual labor and prioritizing critical cases, AI helps save time and resources for medical practitioners, ultimately leading to increased productivity and improved patient care.

## Risks of AI in Health Care

The risks of AI in health care are listed in Textbox 2.

**Textbox 2.** Risks of artificial intelligence (AI) in health care.

**Risks of AI in health care**

- AI diagnosis is not always superior to human diagnosis
- AI programs may be difficult to understand and overly ambitious
- Implementation issues
- Transparency issues and risks with data sharing
- Biases
- Mistakes in disease diagnosis or AI cannot be held accountable
- Data availability and accessibility
- Regulatory concerns
- Social challenges

### *AI Diagnosis Is Not Always Superior to Human Diagnosis.*

While AI has the potential to improve accurate diagnosis, it is not always superior to human diagnosis. Early AI systems, such as the MYCIN program developed in the 1970s, showed promise in diagnosing and treating diseases but did not outperform human diagnosticians [39]. These RBES needed better integration with clinical workflows and medical record systems to be practical and effective. In addition, AI models can suffer from overfitting, generating irrelevant correlations between patient characteristics and outcomes, which can lead to incorrect predictions when applied to new cases [40].

### *Challenges in Understanding and Ambition of AI Programs*

Physicians may find it challenging to understand AI programs, particularly in complex domains like cancer diagnosis and treatment. IBM's Watson program, which combines ML and NLP, garnered attention for its focus on precision medicine. However, integrating Watson into care processes and systems and programming it to handle certain types of cancer has proven difficult [41]. The ambition of AI programs, such as tackling complex cancer therapy, may exceed their current capabilities.

### *Implementation Issues*

Implementing AI in health care faces several challenges. RBES embedded in electronic health care systems are commonly used but may lack the accuracy of algorithmic systems based on ML. These RBES struggle to keep up with evolving medical knowledge and handle large amounts of data [42]. The lack of empirical evidence confirming the efficacy of AI-based treatments in prospective clinical trials hinders successful implementation [43]. Much of the AI research in health care is preclinical and lacks real-world validation [44]. Integration into physician workflow is crucial for successful implementation, but there are limited examples of AI integration into clinical

treatment, and training physicians to use AI effectively can be a time-consuming process [45].

### *Transparency Issues and Risks With Data Sharing*

The use of intelligent machines in health care decision-making raises concerns about accountability, transparency, permission, and privacy [2]. Understanding and interpreting AI systems, such as deep learning algorithms used in image analysis, can be challenging [2]. Physicians who lack comprehension of the inner workings of AI models may struggle to communicate the medical treatment process to patients [46]. Increased reliance on AI may lead to automated decision-making, limiting the contact and communication between health care workers and patients [46].

The rapid emergence of new technologies in health care has sparked skepticism due to the risks associated with data sharing [17]. There is a need for public norms that ensure data governance and openness, as well as improve patient understanding of how and why data are used [17]. Concerns about privacy violations arise from the collection of large data sets and the potential for AI to anticipate personal information [47]. Patients may perceive this as a violation of their privacy, especially if the findings are made public to third parties [48].

Respecting patient confidentiality and acquiring informed consent for data use are ethically required [49]. AI systems should be protected from privacy breaches to prevent psychological and reputational harm to patients [49]. Recent incidents, such as the misuse of Facebook personal data by Cambridge Analytica and the sharing of patient data without explicit consent by the Royal Free London NHS Foundation Trust, have raised concerns about privacy violations [49,50].

### *Biases*

ML systems in health care can be prone to algorithmic bias, leading to predictions based on noncausal factors like gender or ethnicity [51]. Prejudice and inequality are among the risks associated with health care AI [28]. Biases present in the data used to train AI systems can result in inaccurate outcomes, especially if certain races or genders are underrepresented [28]. Unrepresentative data can further perpetuate health inequities and lead to risk underestimation or overestimation in specific patient populations [52].

### *Mistakes in Disease Diagnosis and Lack of Accountability*

AI systems can make mistakes in patient diagnosis and treatment, creating potential harm [28]. Holding AI systems accountable can be challenging, as liability concerns arise regarding errors and the allocation of responsibility [53]. The lack of explanation from deep learning algorithms can hinder both legal accountability and scientific understanding, potentially eroding patients' trust in the system [54].

Determining accountability for AI failures is an ongoing challenge, as holding the physician accountable may seem unjust, while holding the developer accountable may be too removed from the clinical setting [24]. The question of who should be held accountable when AI systems fail remains to be answered [24].

### *Data Availability and Accessibility*

Large amounts of data from various sources are required to train AI algorithms in health care [55]. However, accessing health data can be challenging due to fragmentation across different platforms and systems [55]. Data availability in health care is limited, and there is often a reluctance to share data between hospitals [56]. The continuous availability of data for ongoing improvement of ML-based systems can be difficult due to organizational resistance [57]. Technological advancements and improved algorithms can help address the problem of limited data sets [57].

### *Regulatory Concerns*

The autolearn feature of AI software poses regulatory challenges as algorithms evolve continuously with use [58]. This creates the need for additional policies and procedures to ensure patient safety [58]. Many countries have yet to formalize regulatory guidelines for assessing AI algorithmic safety, which can hinder AI adoption and lead to risky practices [59]. The lack of industry rules on the ethical usage of AI in health care further complicates the accountability issue [60]. Efforts by the Food and Drug Administration and National Health Service to establish guidelines and standards are ongoing but pose barriers to regulatory approval [60,61].

### *Social Challenges*

Misconceptions about AI replacing health care jobs lead to skepticism and aversion to AI-based interventions [43]. However, the arrival of AI does not necessarily mean job obsolescence but rather job reengineering [62]. Overcoming skepticism and fostering trust in AI requires a better understanding of its capabilities and meaningful public discourse [62]. Improving public and health care professionals' understanding of AI is essential to managing expectations and addressing concerns.

## Discussion

### Principal Findings

This narrative review delves into the dynamic landscape of AI integration in health care, aiming to uncover a spectrum of perspectives, concerns, and opportunities. This exploration encompasses a diverse range of health care settings from different countries and regions, unveiling a rich tapestry of AI adoption. Overall, AI offers tremendous potential and will continue to play a crucial role in future health care decisions. If AI is successfully used, it can reduce pressure on health care workers while improving work quality by lowering mistakes and improving precision. It has the potential to give people more control over their health decisions and can lower avoidable hospitalizations. It can broaden the scope of medical knowledge and build on the present clinical guidelines. Given its advantages and capability to drive the development of precision medicine, it is universally acknowledged to be a much-needed enhancement in medicine. AI is anticipated to eventually master most of the essential domains within health care.

However, there are some difficulties associated with incorporating AI in health care. Acquiring enough data to train

precise algorithms is a continual effort that necessitates a shift in attitude towards data sharing that promotes technical advancement. Specific guidelines on how to securely adopt and evaluate AI technology and research on AI's potential and limitations are required. Robust research is also required to empirically demonstrate the benefits of AI use in the actual world. While the perfect conditions for successful AI adoption may not yet exist, there is still room for AI advancement in health care.

Given that AI has tremendous potential and is the future, there are a few crucial points to consider when using AI in health care. First, given the need for more general agreement in AI governance, it may be impossible to develop AI-based systems whose algorithms can be generalized across all health care domains. As a result, it may be wise to concentrate on systems that can be implemented and used effectively in the health care institutions for which they were designed. Fundamentally, patient care must take priority over the excitement of cutting-edge technology. The AI system's safety and competence must be weighed for use only when appropriate and valuable to patients.

Second, AI in health care must still be complemented with human input. Although AI has advantages in speed and accuracy, physicians are still needed for more cognitively complex or psychological elements and activities. Similarly, although the detection and monitoring of vital disease symptoms are now automated, the objective behind AI is not to eliminate physician input but to focus their expertise on areas where they are more necessary and on what computers cannot and may never imitate. Therefore, focusing on developing complementarity between the use of AI and physicians by training is essential.

In addition, while it is critical to lower expectations, it is also critical not to be excessively gloomy about the role of AI in health care. While physicians may need to comprehend the processes of AI algorithms, most physicians understand magnetic resonance imaging or computed tomography to some level. Despite a lack of individual physician comprehension of their specific process, these studies are extensively used. The lack of transparency in ML algorithms may thus be tolerable if the algorithm's efficacy can be demonstrated. Again, this can be achieved with training and familiarizing the physician with the AI system.

Rather than putting AI to a standard of either perfect or nil results, one should compare the outcomes of using AI to those of the natural world, where physicians can and will make mistakes. Importantly, AI is dynamic in nature and can improve using more extensive data sets. As a result, it is entirely possible that the combined usage of physician and AI input would be more successful over time. However, it is critical not to overstate the status of AI. Its implementation in health care will be a careful, deliberate, and progressive process, including strict control and monitoring of its use and efficacy. AI can help patients and increase the quality of care when combined with input and oversight from health care professionals. AI systems will not wholly replace human clinicians but will supplement their efforts to care for patients. Human therapists may eventually shift towards jobs and job designs that require distinctly human skills, such as enthusiasm and knowledge to use AI in health care.

As global communities live longer lives and the prevalence of chronic disease rises, the rising cost of health care will remain a hot topic among health care stakeholders. It is time to seek the assistance of machines as they can potentially reduce economic costs. Furthermore, coordination between government and private sector industry partners is vital to realize this potential and take advantage of AI's full potential in health care delivery.

With all this, the key challenge for AI in many sectors of health care is ensuring its adoption in daily clinical practice rather than whether the technologies will be equipped to be effective. AI systems must be approved by regulators, linked with electronic health record systems, standardized to the point that similar products perform similarly, taught to medical practitioners, paid for by public or private payer organizations, and modified in the field over time for widespread adoption to occur shortly. Since AI has a significant and lasting impact on lives and is the future of health care, it is essential to address the associated concerns. Given its importance, AI needs proper policy guidelines and regulations regarding its usage in health care to reap its maximum benefits.

## Comparison With Previous Literature

In comparing the findings of this review with existing literature, several key similarities and differences emerge. This review aligns with Gazquez-Garcia et al [63] and Mooghali et al [64] in highlighting the crucial role of health care professional training for effective AI integration. Both emphasize the need for proficiency in AI fundamentals, data analytics, and ethical considerations, reinforcing the notion that successful AI adoption requires a well-prepared workforce. The review also echoes Sapci and Sapci's [65] advocacy for incorporating AI education into medical curricula to address future challenges.

However, this review diverges in its emphasis on practical AI implementation challenges. While Moghadasi et al [66] and Muley et al [67] discuss the risks associated with AI, including the need for enhanced transparency and stakeholder collaboration, this review adds a nuanced perspective on balancing AI's potential benefits with its ethical risks. For instance, this review highlights the importance of human oversight and the complementarity of AI with clinician expertise, which aligns with Morley et al [68] and Zhang and Zhang [69] but also offers additional insights into practical implementation issues not fully covered in the previous reviews.

In terms of public perception, this review supports Kerstan [70] and Castagno and Khalifa [71] by acknowledging that trust in AI is influenced by knowledge and transparency. However, it further explores the dynamic interaction between AI's promise and the necessity for rigorous validation and ethical governance, as discussed by Macrae [72] and Tulk Jesso et al [73]. This review underscores that while AI has the potential to revolutionize health care, its integration must be handled with careful consideration of both practical and ethical dimensions

to achieve meaningful improvements in patient care and outcomes.

### Strengths and Limitations

First, the generalizability of the findings may be affected by the inherent variations in study methodologies, AI implementations, and health care settings across different regions. This heterogeneity introduces variability that could influence the applicability of the conclusions. To mitigate this limitation, rigorous search strategies were used across multiple databases to include a diverse range of studies. Future reviews could benefit from incorporating more standardized studies to enhance generalizability. Second, the reliance on published literature from electronic databases introduces potential publication bias. Studies with positive outcomes related to AI integration in health care may be more likely to be published, which could skew perceptions of AI effectiveness and adoption rates. Efforts were made to address this bias by including a broad range of databases and emphasizing recent literature. Future research should aim to include unpublished studies or grey literature to provide a more balanced view. Third, the rapid evolution of AI technologies means that newer developments and implementations may not have been fully captured in this review. The review focused on the most current literature available at the time of the search to address this issue. Regular updates will be necessary to incorporate the latest advancements and ensure the review remains relevant. In addition, the absence of details around stakeholder engagement could have enriched the study by providing additional depth and perspective. Engaging stakeholders in such a dynamic field would offer diverse viewpoints and further validate the conclusions. Future research should consider incorporating stakeholder engagement to enhance the robustness and applicability of the findings.

Despite these limitations, this review offers several notable strengths. It provides a comprehensive overview of AI integration in health care, leveraging rigorous search strategies across multiple databases to ensure a diverse and current collection of literature. This approach contributes to a nuanced understanding of AI's potential and limitations. Furthermore, the emphasis on recent developments helps ensure that the review reflects the most current trends and advancements in AI technologies.

### Future Directions

Moving forward, further research in the field of AI integration in health care should address several key areas to advance understanding and application. First, studies should prioritize incorporating stakeholder engagement, including health care providers, patients, policymakers, and technology developers, to provide diverse perspectives on AI adoption and implementation strategies, enhancing relevance and acceptance in clinical practice. Second, longitudinal studies are crucial to assess the long-term impacts of AI technologies in health care settings, providing insights into sustainability, scalability, and real-world effectiveness over time. Third, comprehensive research focusing on the ethical implications of AI, including data privacy, algorithm bias, patient consent, and regulatory frameworks, is needed to build trust and ensure responsible deployment. In addition, comparative effectiveness research comparing AI-assisted interventions with standard care protocols can provide evidence of AI's impact on clinical outcomes, patient safety, and health care efficiency. Interdisciplinary collaboration between computer scientists, health care professionals, social scientists, and ethicists is essential to foster innovative approaches aligned with health care needs. Education and training programs for health care professionals on AI technologies will ensure proficiency in interpreting AI-generated insights and integrating them into patient care effectively. Finally, research should explore how AI can reduce health care disparities and improve access to quality care, particularly in underserved communities and low-resource settings. Addressing these priorities will realize AI's potential in transforming health care delivery and improving patient outcomes globally.

### Conclusions

In summary, AI presents a transformative force in health care, with the potential to enhance patient care, reduce errors, and broaden medical knowledge. However, its successful integration requires adaptability; complementarity with human expertise; transparency; and a deliberate, incremental approach. AI's impact on health care is evolutionary, not revolutionary, and collaboration between stakeholders, standardization, education, and robust policies are essential to harness its full potential while upholding patient-centric care and innovation.

## Data Availability

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

## Authors' Contributions

As the sole author of this manuscript, MC was responsible for all aspects of the study, including conceptualization, literature review, writing, and editing.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategy across databases for artificial intelligence (AI) integration in health care.
[DOCX File , 14 KB - ijmr_v13i1e53616_app1.docx ]

## References


1. Ghafur S, van Dael J, Leis M, Darzi A, Sheikh A. Public perceptions on data sharing: key insights from the UK and the USA. Lancet Digit Health 2020;2(9):e444-e446 [FREE Full text] [doi: 10.1016/S2589-7500(20)30161-8] [Medline: 32838250]
2. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
3. Lee SI, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat Commun 2018;9(1):42 [FREE Full text] [doi: 10.1038/s41467-017-02465-5] [Medline: 29298978]
4. Sordo M. Introduction to Neural Networks in Healthcare. In: Open Clinical Knowledge Management for Medical Care. Boston, MA: Margarita Sordo; 2002.
5. Fakoor R, Ladhak F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. 2013 Presented at: The 30th International Conference on Machine Learning (ICML 2013), WHEALTH Workshop; June 16-21, 2023; Atlanta, GA URL: https://www.researchgate.net/publication/281857285_Using_deep_learning_to_enhance_cancer_diagnosis_and_classification
6. Vial A, Stirling D, Field M, Ros M, Ritz C, Carolan M, et al. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review. Transl Cancer Res 2018;7(3):803-816. [doi: 10.21037/tcr.2018.05.02]
7. Davenport TH, Glaser J. Just-in-time delivery comes to knowledge management. Harv Bus Rev 2002;80(7):107-11, 126. [Medline: 12140850]
8. Quinn TP, Senadeera M, Jacobs S, Coghlan S, Le V. Trust and medical AI: the challenges we face and the expertise needed to overcome them. J Am Med Inform Assoc 2021;28(4):890-894 [FREE Full text] [doi: 10.1093/jamia/ocaa268] [Medline: 33340404]
9. Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. Information Fusion 2023;96:156-191. [doi: 10.1016/j.inffus.2023.03.008]
10. Wesson P, Hswen Y, Valdes G, Stojanovski K, Handley MA. Risks and opportunities to ensure equity in the application of big data research in public health. Annu Rev Public Health 2022;43:59-78 [FREE Full text] [doi: 10.1146/annurev-publhealth-051920-110928] [Medline: 34871504]
11. Powezka K, Slater L, Wall M, Gkoutos G, Juszczak M. Source of data for artificial intelligence applications in vascular surgery - a scoping review. medRxiv Preprint posted online on October 4, 2023. [doi: 10.1101/2023.10.03.23296506]
12. Gaviria-Valencia S, Murphy SP, Kaggal VC, McBane Ii RD, Rooke TW, Chaudhry R, et al. Near real-time natural language processing for the extraction of abdominal aortic aneurysm diagnoses from radiology reports: algorithm development and validation study. JMIR Med Inform 2023;11:e40964 [FREE Full text] [doi: 10.2196/40964] [Medline: 36826984]
13. Liu W, Zhang X, Lv H, Li J, Liu Y, Yang Z, et al. Using a classification model for determining the value of liver radiological reports of patients with colorectal cancer. Front Oncol 2022;12:913806 [FREE Full text] [doi: 10.3389/fonc.2022.913806] [Medline: 36479085]
14. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. NPJ Digit Med 2022;5(1):194 [FREE Full text] [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
15. Bhayana R. Chatbots and large language models in radiology: a practical primer for clinical and research applications. Radiology 2024;310(1):e232756. [doi: 10.1148/radiol.232756] [Medline: 38226883]
16. Hu Y, Chen Q, Du J, Peng X, Keloth VK, Zuo X, et al. Improving large language models for clinical named entity recognition via prompt engineering. J Am Med Inform Assoc 2024;31(9):1812-1820. [doi: 10.1093/jamia/ocad259] [Medline: 38281112]
17. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. Lancet Digit Health 2021;3(3):e195-e203 [FREE Full text] [doi: 10.1016/S2589-7500(20)30292-2] [Medline: 33478929]
18. Rysavy M. Evidence-based medicine: a science of uncertainty and an art of probability. Virtual Mentor 2013;15(1):4-8 [FREE Full text] [doi: 10.1001/virtualmentor.2013.15.1.fred1-1301] [Medline: 23356799]
19. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
20. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res 2017;4(1):e000234 [FREE Full text] [doi: 10.1136/bmjresp-2017-000234] [Medline: 29435343]
21. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Rep 2016;6:24454 [FREE Full text] [doi: 10.1038/srep24454] [Medline: 27079888]
22. Young SD, Crowley JS, Vermund SH. Artificial intelligence and sexual health in the USA. Lancet Digit Health 2021;3(8):e467-e468 [FREE Full text] [doi: 10.1016/S2589-7500(21)00117-5] [Medline: 34325852]
23. Madsen LB. Data-Driven Healthcare: How Analytics and BI are Transforming The Industry. Hoboken, NJ: Wiley Publishing; 2014.

24. Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. Br Med Bull 2021;139(1):4-15. [doi: 10.1093/bmb/ldab016] [Medline: 34405854]
25. Jones LD, Golan D, Hanna S, Ramachandran M. Artificial intelligence, machine learning and the evolution of healthcare: A bright future or cause for concern? Bone and Joint Research 2018;7(3):223-225 [FREE Full text]
26. Hashimoto DA, Ward TM, Meireles OR. The role of artificial intelligence in surgery. Adv Surg 2020;54:89-101. [doi: 10.1016/j.yasu.2020.05.010] [Medline: 32713441]
27. Akbar Safav A, Fekri P, Setoodeh P, Zadeh MH. Toward deep secure tele-surgery system. 2018 Presented at: The 16th International Conference on Scientific Computing (CSC'18); October 13, 2024; Las Vegas, NV URL: https://www.researchgate.net/publication/346502758_Toward_Deep_Secure_Tele-surgery_System
28. Shaheen MY. AI in healthcare: medical and socioeconomic benefits and challenges. OSF Preprint posted online on September 28, 2021. [doi: 10.31219/osf.io/um38t]
29. Luxton DD. An introduction to artificial intelligence in behavioral and mental health care. In: Artificial Intelligence in Behavioral and Mental Health Care. London, United Kingdom: Elsevier Academic Press; 2016:1-26.
30. Lovejoy CA, Buch V, Maruthappu M. Technology and mental health: the role of artificial intelligence. Eur Psychiatry 2019 Jan;55:1-3. [doi: 10.1016/j.eurpsy.2018.08.004] [Medline: 30384105]
31. Habermann J. Language and psycho-social well-being. Knowledge Common Works Preprint posted online in 2021. [doi: 10.17613/gb58-rx71]
32. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
33. Nguyen LT, Do TTH. Artificial Intelligence in Healthcare: A New Technology Benefit for Patients and Doctors. 2019 Presented at: Proceedings of the Portland International Conference on Management of Engineering and Technology (PICMET); August 25-29, 2019; Portland, OR p. 1-15. [doi: 10.23919/picmet.2019.8893884]
34. Beck JT, Rammage M, Jackson GP, Preininger AM, Dankwa-Mullan I, Roebuck MC, et al. Artificial intelligence tool for optimizing eligibility screening for clinical trials in a large community cancer center. JCO Clin Cancer Inform 2020;4:50-59 [FREE Full text] [doi: 10.1200/CCI.19.00079] [Medline: 31977254]
35. Ichikawa D, Saito T, Ujita W, Oyama H. How can machine-learning methods assist in virtual screening for hyperuricemia? A healthcare machine-learning approach. J Biomed Inform 2016;64:20-24 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.012] [Medline: 27658886]
36. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020;368:l6927 [FREE Full text] [doi: 10.1136/bmj.l6927] [Medline: 32198138]
37. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. JAMA 2018;319(1):19-20. [doi: 10.1001/jama.2017.19198] [Medline: 29261830]
38. Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep 2014;16(1):441. [doi: 10.1007/s11886-013-0441-8] [Medline: 24338557]
39. Bush J. How AI is taking the scut work out of health care. Harvard Business Review. 2018 Mar 5. URL: https://hbr.org/2018/03/how-ai-is-taking-the-scut-work-out-of-health-care [accessed 2024-10-04]
40. Wiens J, Shenoy ES. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. Clin Infect Dis 2018;66(1):149-153 [FREE Full text] [doi: 10.1093/cid/cix731] [Medline: 29020316]
41. Swetlitz L, Ross C. IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. STAT. 2017 Sep 5. URL: https://www.statnews.com/2017/09/05/watson-ibm-cancer/ [accessed 2017-09-05]
42. Davenport TH. The AI Advantage: How to Put the Artificial Intelligence Revolution to Work. Cambridge, MA: The MIT Press; 2018.
43. Sun TQ, Medaglia R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Government Information Quarterly 2019;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
44. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med 2018;1:5 [FREE Full text] [doi: 10.1038/s41746-017-0012-2] [Medline: 31304291]
45. Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine. Emerg Med Australas 2018;30(6):870-874. [doi: 10.1111/1742-6723.13145] [Medline: 30014578]
46. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med 2018;15(11):e1002689. [doi: 10.1371/journal.pmed.1002689]
47. Marwan M, Kartit A, Ouahmane H. Security enhancement in healthcare cloud using machine learning. Procedia Computer Science 2018;127:388-397. [doi: 10.1016/j.procs.2018.01.136]
48. van der Schaar M, Alaa AM, Floto A, Gimson A, Scholtes S, Wood A, et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. Mach Learn 2021;110(1):1-14 [FREE Full text] [doi: 10.1007/s10994-020-05928-x] [Medline: 33318723]

49. Dawson D, Schleiger E, McLaughlin J, Robinson C, Quezada G, Scowcroft J, et al. Artificial Intelligence Australia's Ethics Framework A Discussion Paper. Eveleigh, Australia: Data 61 CSIRO; 2019.
50. Powles J, Hodson H. Google deepMind and healthcare in an age of algorithms. Health Technol (Berl) 2017;7(4):351-367 [FREE Full text] [doi: 10.1007/s12553-017-0179-1] [Medline: 29308344]
51. Davenport TH, Dreyer KJD. AI will change radiology, but it won't replace radiologists. Harvard Business Publishing Education. 2018. URL: https://hbsp.harvard.edu/search?action=&author=Keith+J+Dreyer+DO&activeTab=products [accessed 2024-10-08]
52. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. In: Martin K, editor. Ethics of Data and Analytics. New York, NY: Auerbach Publications; 2022.
53. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med 2019;112(1):22-28 [FREE Full text] [doi: 10.1177/0141076818815510] [Medline: 30507284]
54. Wang F, Preininger A. AI in health: state of the art, challenges, and future directions. Yearb Med Inform 2019;28(1):16-26 [FREE Full text] [doi: 10.1055/s-0039-1677908] [Medline: 31419814]
55. Jianying Hu APFW. Data-driven analytics for personalized healthcare. Healthcare Information Management Systems 2016:529-554 [FREE Full text]
56. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. J Am Coll Cardiol 2018;71(23):2668-2679 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.521] [Medline: 29880128]
57. Lopez K, Fodeh SJ, Allam A, Brandt CA, Krauthammer M. Reducing annotation burden through multimodal learning. Front Big Data 2020;3:19 [FREE Full text] [doi: 10.3389/fdata.2020.00019] [Medline: 33693393]
58. Deciding when to submit a 510(k) for a software change to an existing device. Food and Drug Administration. 2017 Oct. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/deciding-when-submit-510k-software-change-existing-device [accessed 2024-11-06]
59. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. Science 2019;363(6429):810-812. [doi: 10.1126/science.aaw0029]
60. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
61. Ramesh AN, Kambhampati C, Monson JRT, Drew PJ. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004;86(5):334-338 [FREE Full text] [doi: 10.1308/147870804290] [Medline: 15333167]
62. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
63. Gazquez-Garcia J, Sánchez-Bocanegra CL, Sevillano JL. Artificial intelligence in the health sector: key skills for future health professionals. JMIR Preprints Preprint posted online on March 7, 2024. [doi: 10.2196/preprints.58161]
64. Mooghali M, Stroud AM, Yoo DW, Barry BA, Grimshaw AA, Ross JS, et al. Barriers and facilitators to trustworthy and ethical AI-enabled medical care from patient's and healthcare provider's perspectives: a literature review. medRxiv Preprint posted on online on October 2, 2023. [doi: 10.1101/2023.10.02.23296447]
65. Sapci AH, Sapci HA. Artificial intelligence education and tools for medical and health informatics students: systematic review. JMIR Med Educ 2020;6(1):e19285 [FREE Full text] [doi: 10.2196/19285] [Medline: 32602844]
66. Moghadasi N, Valdez RS, Piran M, Moghaddasi N, Linkov I, Polmateer TL, et al. Risk analysis of artificial intelligence in medicine with a multilayer concept of system order. Systems 2024;12(2):47. [doi: 10.3390/systems12020047]
67. Muley A, Muzumdar P, Kurian G, Basyal GP. Risk of AI in healthcare: a comprehensive literature review and study framework. AJMAH 2023;21(10):276-291. [doi: 10.9734/ajmah/2023/v21i10903] [Medline: 39250570]
68. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. SSRN Journal 2020. [doi: 10.2139/ssrn.3830408]
69. Zhang J, Zhang ZM. Ethics and governance of trustworthy medical artificial intelligence. BMC Med Inform Decis Mak 2023;23(1):7-7. [doi: 10.1186/s12911-023-02103-9]
70. Kerstan S, Bienefeld N, Grote G. Choosing human over AI doctors? how comparative trust associations and knowledge relate to risk and benefit perceptions of AI in healthcare. Risk Anal 2024;44(4):939-957. [doi: 10.1111/risa.14216] [Medline: 37722964]
71. Castagno S, Khalifa M. Perceptions of artificial intelligence among healthcare staff: a qualitative survey study. Front Artif Intell 2020;3:578983 [FREE Full text] [doi: 10.3389/frai.2020.578983] [Medline: 33733219]
72. Macrae C. Governing the safety of artificial intelligence in healthcare. BMJ Qual Saf 2019;28(6):495-498. [doi: 10.1136/bmjqs-2019-009484] [Medline: 30979783]
73. Tulk Jesso S, Kelliher A, Sanghavi H, Martin T, Henrickson Parker S. Inclusion of clinicians in the development and evaluation of clinical artificial intelligence tools: a systematic literature review. Front Psychol 2022;13:830345 [FREE Full text] [doi: 10.3389/fpsyg.2022.830345] [Medline: 35465567]


## Abbreviations

**AI:** artificial intelligence

**BERT:** Bidirectional Encoder Representations from Transformers
**GPT:** Generative Pre-trained Transformer
**MeSH:** Medical Subject Headings
**ML:** machine learning
**NLP:** natural language processing
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RBES:** rule-based expert systems
**STI:** sexually transmitted infection

*Edited by T de Azevedo Cardoso; submitted 12.10.23; peer-reviewed by J Walsh, B Delaney, W Yang; comments to author 19.03.24; revised version received 17.06.24; accepted 19.09.24; published 18.11.24.*

*Please cite as:*
*Chustecki M*
*Benefits and Risks of AI in Health Care: Narrative Review*
*Interact J Med Res 2024;13:e53616*
*URL: https://www.i-jmr.org/2024/1/e53616*
*doi:10.2196/53616*
*PMID:*

Original Paper

# The Effect of Body Temperature Changes on the Course of Treatment in Patients With Pneumonia and Sepsis: Results of an Observational Study

Domen Guzelj[1*]; Anže Grubelnik[1*]; Nina Greif[1*]; Petra Povalej Bržan[2*], PhD; Jure Fluher[3*], MD; Žiga Kalamar[3*], MD; Andrej Markota[3*], MD, PhD

[1]Faculty of Medicine, University of Maribor, Maribor, Slovenia

[2]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[3]Medical 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 $FiO_2$ 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 1.1 g vs mean 0.4, 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.

***(Interact J Med Res 2024;13:e52590)*** doi:10.2196/52590

**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.

### Measurements

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 (Rüsch 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-N 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 $FiO_2$ 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*=.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).

| Label and variable | Group | | | Total | *P* value |
|---|---|---|---|---|---|
| | P-N[a] | P-P[b] | P-HAP[c] | | |
| Patients, n (%) | 25 (21.4) | 50 (42.7) | 42 (35.9) | 117 (100) | |
| **Gender, n (%)** | | | | | .26 |
| Female | 5 (15.2) | 18 (54.5) | 10 (30.3) | 33 (28.2) | |
| Male | 20 (23.8) | 32 (38.1) | 32 (38.1) | 84 (71.8) | |
| **Age (years)** | | | | | .21 |
| Median (IQR) | 62.0 (52.0-74.0) | 64.0 (57.2-71.8) | 67.0 (62.0-76.0) | 65.0 (57.0-74.0) | |
| Min-max | 23.0-84.0 | 26.0-84.0 | 39.0-86.0 | 23.0-86.0 | |
| **BMI (kg/m$^2$)** | | | | | .19 |
| Mean (SD) | 28.7 (7.0) | 27.6 (4.3) | 29.7 (6.4) | 28.6 (5.8) | |
| Min-max | 17.2-47.7 | 18.8-39.2 | 16.3-44.8 | 16.3-47.7 | |
| **Length of hospitalization (days)** | | | | | .35 |
| Median (IQR) | 22.0 (17.0-29.0) | 19.5 (15.0-26.8) | 17.5 (13.0-26.8) | 20.0 (14.0-26.9) | |
| Min-max | 8.0-53.0 | 10.0-43.0 | 8.0-69.0 | 8.0-69.0 | |
| **ICU survival, n (%)** | | | | | .26 |
| Died | 6 (15.0) | 16 (40.0) | 18 (45.0) | 40 (34.2) | |
| Survived | 19 (26.1) | 34 (44.2) | 24 (31.2) | 77 (65.8) | |
| **Hospital survival, n (%)** | | | | | .59 |
| Died | 13 (18.3) | 31 (43.7) | 27 (38.0) | 71 (60.7) | |
| Survived | 12 (26.1) | 19 (41.3) | 15 (32.6) | 46 (39.3) | |
| **SOFA[d] at the time of admission** | | | | | .76 |
| Median (IQR) | 9.5 (7.0-12.0) | 10.0 (9.0-12.0) | 10.0 (9.0-12.0) | 10.0 (9.0-12.0) | |
| Min-max | 5.0-16.0 | 5.0-15.0 | 4.0-15.0 | 4.0-16.0 | |
| **APACHE[e] II at the time of admission** | | | | | .53 |
| Mean (SD) | 20.4 (6.3) | 22.0 (6.4) | 22.1 (7.2) | 21.7 (6.6) | |
| Min-max | 6.0-32.0 | 7.0-36.0 | 5.0-40.0 | 5.0-40.0 | |
| Change of antibiotic in 7 days after admission (no), n (%) | 7 (18.4) | 17 (44.7) | 14 (36.8) | 38 (32.5) | .86 |

[a]P-N: patients with sterile week 2 tracheal aspirates.

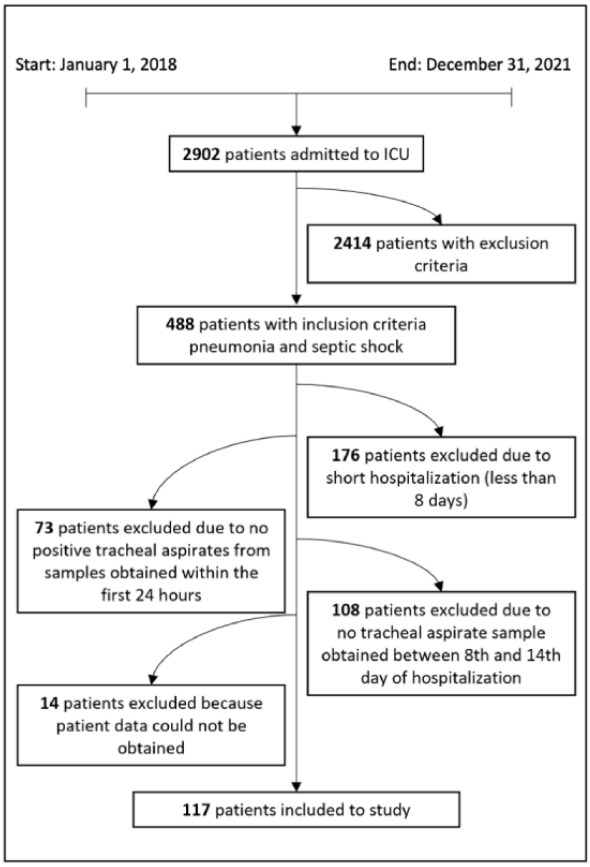
[b]P-HAP: patients with a different pathogen in week 2 tracheal aspirates.

[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.

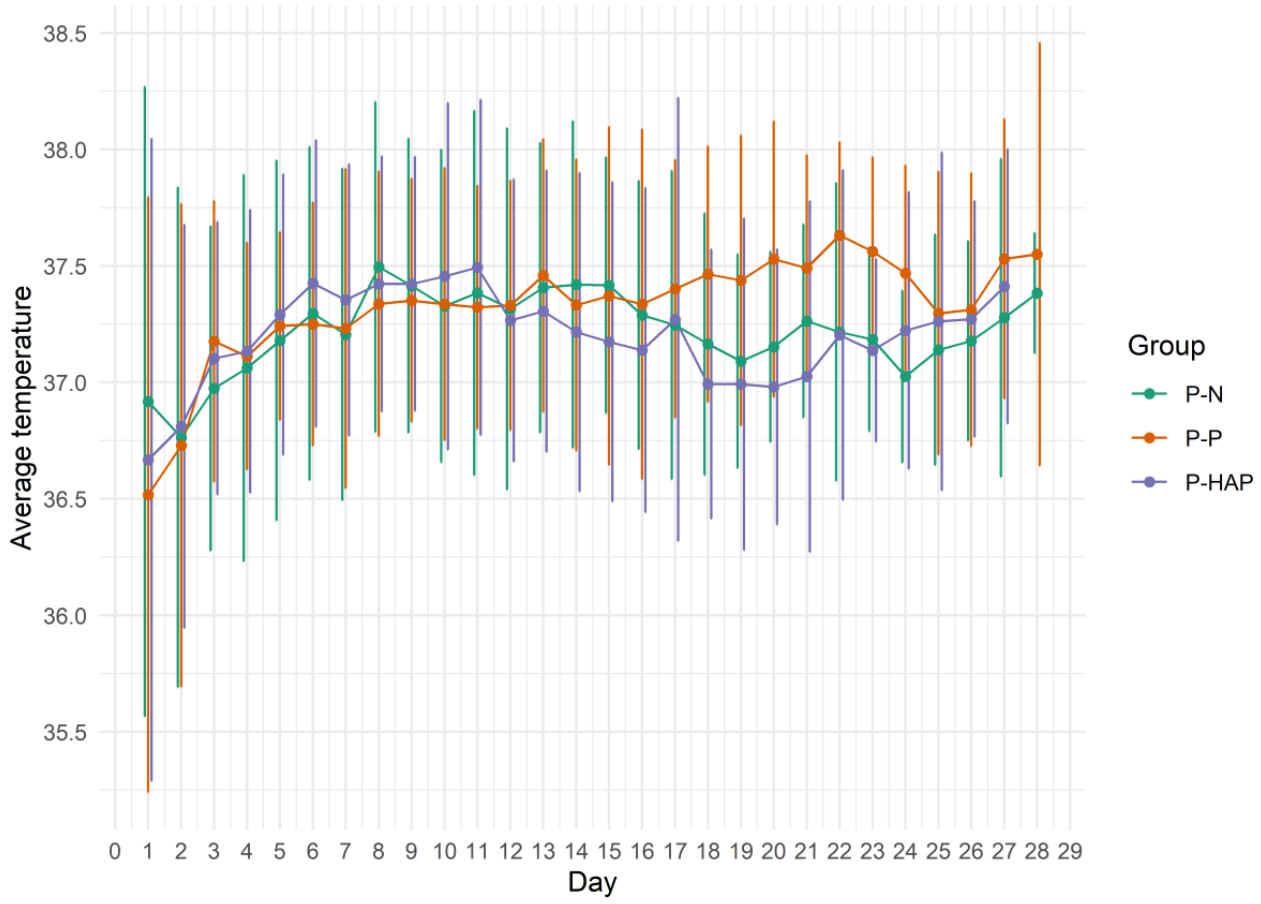
**Figure 1.** Study patient population and inclusion flowchart. ICU: intensive care unit.




## 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 (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 $FiO_2$ 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).

**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.

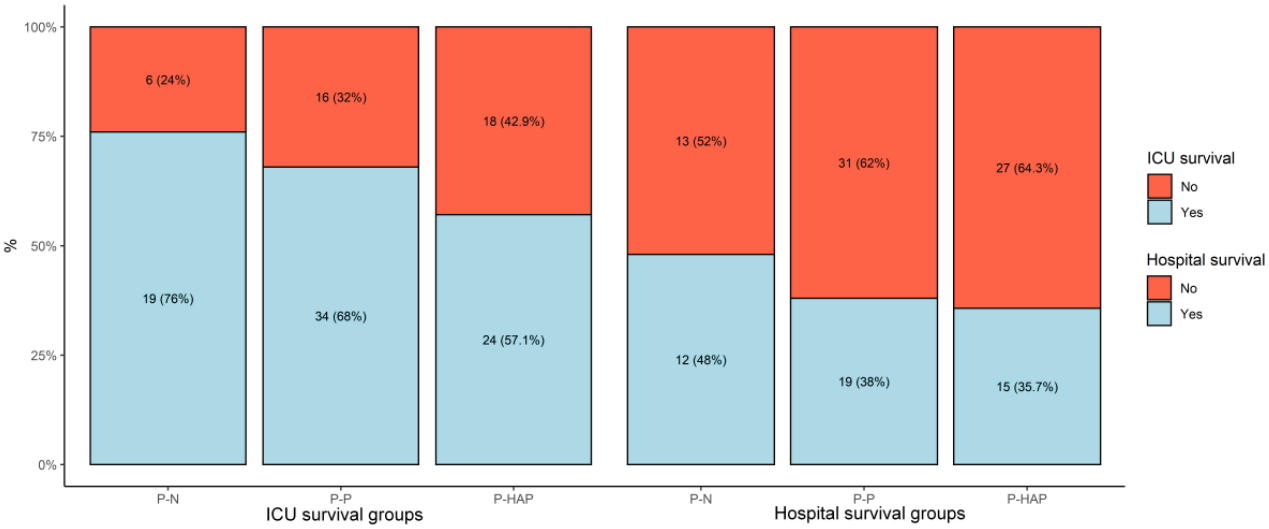
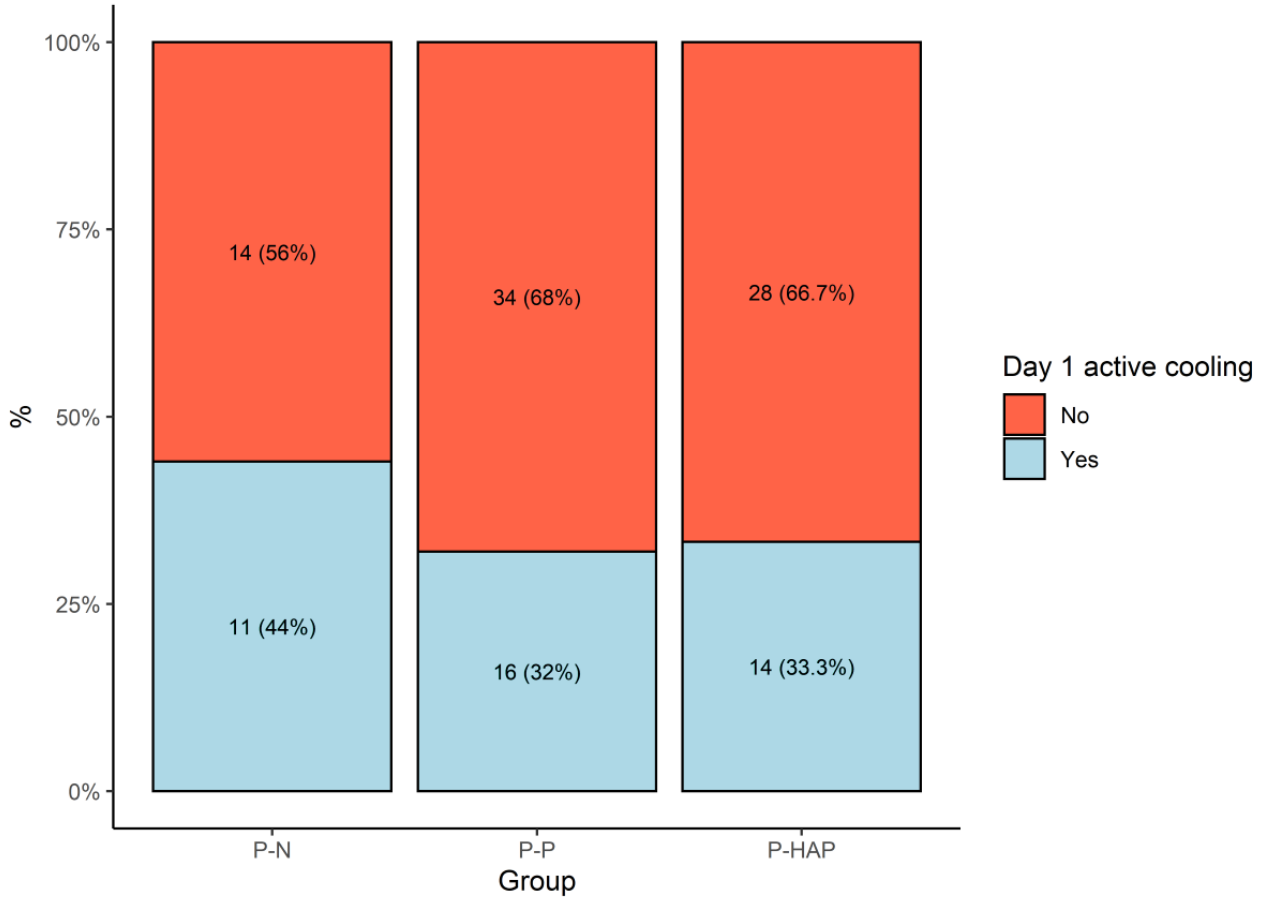


**Figure 4.** Active cooling on day one of hospitalization in the intensive care unit divided by groups. P-HAP: patients with a different pathogen in week 2 tracheal aspirates; P-N: patients with sterile week 2 tracheal aspirates; P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates.




## 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.

| Genus | Group, n (%) | | | Total, n (%) |
|---|---|---|---|---|
| | P-P[a] | P-HAP[b] | P-N[c] | |
| **Gram-positive bacteria** | | | | |
| **Sample 1** | | | | |
| *Staphylococcus* | 8 (4.3) | 15 (8.0) | 6 (3.2) | 29 (15.4) |
| *Streptococcus* | 5 (2.7) | 11 (5.9) | 5 (2.7) | 21 (11.2) |
| *Corynebacterium* | 3 (1.6) | 2 (1.1) | 1 (0.5) | 6 (3.2) |
| Total | 16 (8.5) | 28 (14.9) | 12 (6.4) | 56 (29.8) |
| **Sample 2** | | | | |
| *Staphylococcus* | 5 (3.4) | 2 (1.4) | N/A[d] | 7 (4.7) |
| *Corynebacterium* | 1 (0.7) | 3 (2.0) | N/A | 4 (2.7) |
| *Streptococcus* | 1 (0.7) | 1 (0.7) | N/A | 2 (1.4) |
| *Enterococcus* | 0 (0) | 1 (0.7) | N/A | 1 (0.7) |
| Total | 7 (4.7) | 7 (4.7) | N/A | 14 (9.5) |
| **Gram-negative bacteria** | | | | |
| **Sample 1** | | | | |
| *Klebsiella* | 17 (9.0) | 6 (3.2) | 8 (4.3) | 31 (16.5) |
| *Haemophilus* | 4 (2.1) | 7 (3.7) | 6 (3.2) | 17 (9.0) |
| *Escherichia* | 9 (4.8) | 4 (2.1) | 1 (0.5) | 14 (7.4) |
| *Enterobacter* | 9 (4.8) | 3 (1.6) | 2 (1.1) | 14 (7.4) |
| *Pseudomonas* | 10 (5.3) | 0 (0) | 1 (0.5) | 11 (5.9) |
| Other | 23 (12.2) | 7 (3.7) | 7 (3.7) | 37 (19.7) |
| Total | 72 (38.3) | 27 (14.4) | 25 (13.3) | 124 (66) |
| **Sample 2** | | | | |
| *Klebsiella* | 16 (10.8) | 12 (8.1) | N/A | 28 (18.9) |
| *Pseudomonas* | 16 (10.8) | 8 (5.4) | N/A | 24 (16.2) |
| *Escherichia* | 11 (7.4) | 4 (2.7) | N/A | 15 (10.1) |
| *Enterobacter* | 8 (5.4) | 6 (4.1) | N/A | 14 (9.5) |
| *Acinetobacter* | 10 (6.8) | 3 (2.0) | N/A | 13 (8.8) |
| Other | 18 (12.2) | 16 (10.8) | N/A | 34 (23.0) |
| Total | 79 (53.4) | 49 (33.1) | N/A | 128 (86.5) |
| **Fungi** | | | | |
| **Sample 1** | | | | |
| *Candida* | 1 (0.5) | 6 (3.2) | 1 (0.5) | 8 (4.3) |
| Total | 89 (47.3) | 61 (32.4) | 38 (20.2) | 188 (100) |
| **Sample 2** | | | | |
| *Candida* | 5 (3.4) | 1 (0.7) | N/A | 6 (4.1) |
| Total | 91 (61.5) | 57 (38.5) | N/A | 148 (100) |

[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. Rumbus 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

1. Walter EJ, Hanna-Jumma S, Carraretto M, Forni L. The pathophysiological basis and consequences of fever. Crit Care 2016 Jul 14;20(1):200 [FREE Full text] [doi: 10.1186/s13054-016-1375-5] [Medline: 27411542]
2. Owens CA. The value of fever therapy for gonorrhea. JAMA 1936 Dec 12;107(24):1942. [doi: 10.1001/jama.1936.02770500008003]
3. Wagner-Jaureggj J. The history of the malaria treatment of general paralysis. Am J Psychiatry 1946 Mar;102:577-582. [doi: 10.1176/ajp.102.5.577] [Medline: 21026839]
4. Drewry A, Mohr N, Ablordeppey E, Dalton C, Doctor R, Fuller B, et al. Therapeutic hyperthermia is associated with improved survival in afebrile critically ill patients with sepsis: a pilot randomized trial. Crit Care Med 2022 Jun 01;50(6):924-934 [FREE Full text] [doi: 10.1097/CCM.0000000000005470] [Medline: 35120040]
5. Evans SS, Repasky EA, Fisher DT. Fever and the thermal regulation of immunity: the immune system feels the heat. Nat Rev Immunol 2015 Jun;15(6):335-349 [FREE Full text] [doi: 10.1038/nri3843] [Medline: 25976513]
6. Markota A, Skok K, Kalamar Ž, Fluher J, Gorenjak M. Better control of body temperature is not associated with improved hemodynamic and respiratory parameters in mechanically ventilated patients with sepsis. J Clin Med 2022 Feb 24;11(5):1211 [FREE Full text] [doi: 10.3390/jcm11051211] [Medline: 35268302]
7. Laupland K, Shahpori R, Kirkpatrick A, Ross T, Gregson D, Stelfox H. Occurrence and outcome of fever in critically ill adults. Crit Care Med 2008 May;36(5):1531-1535. [doi: 10.1097/CCM.0b013e318170efd3] [Medline: 18434882]
8. Doyle J, Schortgen F. Should we treat pyrexia? And how do we do it? Crit Care 2016 Oct 03;20(1):303 [FREE Full text] [doi: 10.1186/s13054-016-1467-2] [Medline: 27716372]
9. Gursel G, Demirtas S. Value of APACHE II, SOFA and CPIS scores in predicting prognosis in patients with ventilator-associated pneumonia. Respiration 2006;73(4):503-508. [doi: 10.1159/000088708] [Medline: 16205047]
10. Depuydt P, Vandijck D, Bekaert M, Decruyenaere J, Blot S, Vogelaers D, et al. Determinants and impact of multidrug antibiotic resistance in pathogens causing ventilator-associated-pneumonia. Crit Care 2008;12(6):R142 [FREE Full text] [doi: 10.1186/cc7119] [Medline: 19014695]
11. Qiao Q, Lu G, Li M, Shen Y, Xu D. Prediction of outcome in critically ill elderly patients using APACHE II and SOFA scores. J Int Med Res 2012;40(3):1114-1121 [FREE Full text] [doi: 10.1177/147323001204000331] [Medline: 22906285]
12. Kushimoto S, Gando S, Saitoh D, Mayumi T, Ogura H, Fujishima S, JAAM Sepsis Registry Study Group. The impact of body temperature abnormalities on the disease severity and outcome in patients with severe sepsis: an analysis from a multicenter, prospective survey of severe sepsis. Crit Care 2013 Nov 13;17(6):R271 [FREE Full text] [doi: 10.1186/cc13106] [Medline: 24220071]
13. Shimazui T, Nakada T, Walley K, Oshima T, Abe T, Ogura H, et al. Significance of body temperature in elderly patients with sepsis. Crit Care 2020 Jun 30;24(1):387 [FREE Full text] [doi: 10.1186/s13054-020-02976-6] [Medline: 32605659]
14. Lee B, Inui D, Suh G, Kim J, Kwon J, Park J, et al. Association of body temperature and antipyretic treatments with mortality of critically ill patients with and without sepsis: multi-centered prospective observational study. Crit Care 2012 Feb 28;16(1):R33 [FREE Full text] [doi: 10.1186/cc11211] [Medline: 22373120]

15. Rumbus Z, Matics R, Hegyi P, Zsiboras C, Szabo I, Illes A, et al. Fever is associated with reduced, hypothermia with increased mortality in septic patients: a meta-analysis of clinical trials. PLoS One 2017;12(1):e0170152 [FREE Full text] [doi: 10.1371/journal.pone.0170152] [Medline: 28081244]
16. Thomas-Rüddel DO, Hoffmann P, Schwarzkopf D, Scheer C, Bach F, Komann M, et al. Fever and hypothermia represent two populations of sepsis patients and are associated with outside temperature. Crit Care 2021 Oct 21;25(1):368 [FREE Full text] [doi: 10.1186/s13054-021-03776-2] [Medline: 34674733]
17. Zhang Z, Chen L, Ni H. Antipyretic therapy in critically ill patients with sepsis: an interaction with body temperature. PLoS One 2015;10(3):e0121919 [FREE Full text] [doi: 10.1371/journal.pone.0121919] [Medline: 25822614]
18. Ye S, Xu D, Zhang C, Li M, Zhang Y. Effect of antipyretic therapy on mortality in critically ill patients with sepsis receiving mechanical ventilation treatment. Can Respir J 2017;2017:3087505. [doi: 10.1155/2017/3087505] [Medline: 28386165]
19. Young P, Saxena M, Bellomo R, Freebairn R, Hammond N, van Haren F, et al. Acetaminophen for fever in critically ill patients with suspected infection. N Engl J Med 2015 Dec 03;373(23):2215-2224. [doi: 10.1056/NEJMoa1508375] [Medline: 26436473]
20. Young P, Bellomo R, Bernard G, Niven D, Schortgen F, Saxena M, et al. Fever control in critically ill adults. An individual patient data meta-analysis of randomised controlled trials. Intensive Care Med 2019 Apr;45(4):468-476. [doi: 10.1007/s00134-019-05553-w] [Medline: 30741326]

## 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

*Edited by T de Azevedo Cardoso; submitted 13.09.23; peer-reviewed by S Liu, E Kawamoto; comments to author 03.11.23; revised version received 24.11.23; accepted 06.12.23; published 01.03.24.*

*Please cite as:*
*Guzelj D, Grubelnik A, Greif N, Povalej Bržan P, Fluher J, Kalamar Ž, Markota A*
*The Effect of Body Temperature Changes on the Course of Treatment in Patients With Pneumonia and Sepsis: Results of an Observational Study*
*Interact J Med Res 2024;13:e52590*
*URL: https://www.i-jmr.org/2024/1/e52590*
*doi:10.2196/52590*
*PMID:38427413*

Original Paper

# Predictive Value of Physiological Values and Symptom Scores for Exacerbations in Bronchiectasis and Chronic Obstructive Pulmonary Disease With Frequent Exacerbations: Longitudinal Observational Cohort Study

Thomas Llewelyn Jones[1,2], MA, MBBCHIR, MRCP; Claire Roberts[1], MBChB, MRCP; Scott Elliott[3], BSc; Sharon Glaysher[3], PhD; Ben Green[1], MBBS, DM; Janis K Shute[2†], PhD; Anoop J Chauhan[1,2], MBChB, PhD

[1]Department of Respiratory Medicine, Portsmouth Hospitals University NHS Trust, Portsmouth, United Kingdom

[2]School of Pharmacy and Biomedical Science, University of Portsmouth, Portsmouth, United Kingdom

[3]Translational Research Laboratory, Portsmouth Hospitals University NHS Trust, Portsmouth, United Kingdom

[†]deceased

**Corresponding Author:**
Thomas Llewelyn Jones, MA, MBBCHIR, MRCP
Department of Respiratory Medicine
Portsmouth Hospitals University NHS Trust
Southwick Hill Rd
Portsmouth, PO6 3LY
United Kingdom
Phone: 44 02392286000
Email: thomas.jones@porthosp.nhs.uk

## Abstract

**Background:** COPD (chronic obstructive pulmonary disease) and bronchiectasis are common, and exacerbations contribute to their morbidity and mortality. Predictive factors for the frequency of future exacerbations include previous exacerbation frequency and airway colonization. Earlier treatment of exacerbations is likely to reduce severity.

**Objective:** This study tested the hypothesis that, in a population with bronchiectasis, COPD, or both who have frequent exacerbations and airway colonization, changes in symptom scores or physiological variables within 10 days prior to an exacerbation would allow the prediction of the event.

**Methods:** We performed a 6-month, longitudinal, observational, cohort study among 30 participants with bronchiectasis, COPD, or both; at least 2 exacerbations per year; and colonization with *Pseudomonas aeruginosa* or *Haemophilus influenzae*. Daily symptom and physiological data were collected, comprising pulse rate, blood pressure, oxygen saturation, peak flow rate, step count, weight, and temperature. Exacerbations (defined as the onset of new antibiotic use for respiratory symptoms) were collected, and predictive values for abnormal values in the 10 days prior to an exacerbation were calculated.

**Results:** A total of 30 participants were recruited, collecting a total of 39,534 physiological and 25,334 symptom data points across 5358 participant-days; these included 78 exacerbations across 27 participants, with the remaining 3 participants not having exacerbations within the 6-month observation period. Peak flow rate, oxygen saturation, and weight were significantly different at the point of exacerbation (all $P<.001$), but no significant trends around exacerbation were noted and no clinically beneficial predictive value was found in the overall or individually adjusted model. Symptom scores tended to worsen for 10 days on either side of an exacerbation but were of insufficient magnitude for prediction, with area under the receiver operating characteristic curve values of ranging from 0.4 to 0.6.

**Conclusions:** Within this small cohort with bronchiectasis, COPD, or both and airway colonization, physiological and symptom variables did not show sufficient predictive value for exacerbations to be of clinical utility. The self-management education provided as standard of care may be superior to either of these approaches, but benefit in another or larger cohort cannot be excluded.

**International Registered Report Identifier (IRRID):** RR2-10.2196/resprot.6636

*(Interact J Med Res 2024;13:e44397)* doi:10.2196/44397

**KEYWORDS**

COPD; chronic obstructive pulmonary disease; bronchiectasis; predictive models; airway disease; symptom score

## Introduction

COPD (chronic obstructive pulmonary disease) and bronchiectasis are common causes of morbidity and mortality across the world [1,2]. Much of this morbidity and mortality is associated with exacerbations—acute deteriorations of the disease [3,4]. Exacerbations often require additional treatment, up to and including hospitalization, and are a risk factor for progressive disease [5,6]. Certain factors are known to be associated with increased risk of future exacerbations, including recurrent previous exacerbation [7,8] and airway colonization with organisms such as *Pseudomonas aeruginosa* [9,10].

It is thought that earlier treatment of exacerbations is associated with improved outcomes, including time to recovery and reduced risk of a severe exacerbation [11], but overtreatment risks side effects from the treating medications (eg, corticosteroids or antibiotics). Accurate prediction of exacerbations is therefore required. Predictive factors may be physiological (eg, heart rate or oxygen saturations) or symptomatic (eg, degree of breathlessness or fatigue). The monitoring of physiological variables has shown value in predicting mortality in hospital and prehospital care (eg, National Early Warning Score [12,13]), including mortality in patients with COPD [14]. However, the same has not been consistently demonstrated for other metrics, such as exacerbations of disease, and previous studies focusing on lung function monitoring have not demonstrated predictive benefit [15,16]. Multiple predictors of exacerbation risk exist [17-20], such as the COPD Assessment Test score for COPD and the FACED score for bronchiectasis, but these do not help predict individual exacerbations. In order to predict exacerbations in either physiological or symptom modalities, monitoring must be remote and performed in the home environment.

We conducted a longitudinal, observational, cohort study examining the capability of physiological and symptom variables to predict exacerbations of airway disease in participants with chronic bronchitis with frequent exacerbation and airway colonization with *P aeruginosa* or *Haemophilus influenzae.* We hypothesized that changes in symptom scores or physiological variables within 10 days prior to an exacerbation would allow the prediction of the event.

## Methods

### Study Design and Setting

We conducted a 6-month, blinded, longitudinal, observational, cohort study, recruiting 30 adult participants from secondary care with a diagnosis of COPD, bronchiectasis, or both. Participants were recruited from a secondary care clinic during 2014, and the study ran from September 2014 for 6 months.

### Participants

Participants were required to have had a least 2 exacerbations in the last 12 months, at least 1 of which was within the last 6 months. Participants were required to be colonized by *P aeruginosa* or *H influenzae*, demonstrated by at least 2 cultures at exacerbation in the last 12 months without culture of the other organism. Participants were required to be able to give informed consent, comply with study procedures, and produce at least 5 mL of sputum most days. A full study protocol has been previously published [21], and the participant flow diagram is shown in the Figure S1 in Multimedia Appendix 1.

### Variables

At enrollment, participants provided clinical history, underwent spirometry testing, and completed symptom questionnaires including the St. George's Respiratory Questionnaire (SGRQ). A home visit was conducted when participants were provided with home physiological monitoring equipment [21], specifically a digital peak flow meter, pulse oximeter, physical activity (step) tracker, infrared thermometer, automatic sphygmomanometer, and weighing scales (see Table S1 in Multimedia Appendix 1 for the models of the equipment); they were linked by Bluetooth to an iPad, which was also used collected symptom scores on a 10-point visual analog scale for appetite, breathing, cough, energy, and wellness. These data were transmitted daily to a secure cloud-based database, and the study team were blinded to the data until all participants had completed the study, except for a single unblinded technical observer who ensured that data were being received from each participant. Participants were asked to record whether they had started a course of antibiotics at home (defined as a moderate exacerbation as per guidelines) or whether they were in hospital (defined as a severe exacerbation) [22].

### Outcomes

The outcome under investigation was the accuracy of abnormal values of the physiological and symptom variables listed above to predict a moderate exacerbation of airway disease, as defined by patient-reported initiation of antibiotics for a worsening in respiratory symptomatology. Independent variables were therefore symptom scores (appetite, breathing, cough, energy, and wellness) and physiological values (heart rate, blood pressure, peripheral oxygen saturation, temperature, weight, and daily step count).

### Study Size

To establish an association between predictive markers and the occurrence of exacerbations, the study was designed around the identification of a presymptomatic period. Based on clinical expertise and literature review, we determined that a 10-day window before the onset of an exacerbation is a critical timeframe in which deviations in predictive markers can be most reliably attributed to a forthcoming exacerbation. This period selection aligns with the natural history of exacerbations

as detailed in prior studies. Given the adoption of a 10-day cycle as a unit of observation, we anticipated a minimum of 12 such units per participant across an expected follow-up duration of 4 months. This frequency follows from the inclusion criterion requiring participants to have had at least 2 exacerbations in the previous year, paralleling the exacerbation frequency described by Seemungal et al [23].

For the study to hold clinical relevance, we aimed to achieve a positive predictive value (PPV) of 60% and a negative predictive value (NPV) of 90%. This was based on the study team's experience rather than published data. On the assumption that each participant would experience at least 1 exacerbation during the study period and using a significance threshold of 5% with 90% statistical power to discern the prescribed PPV and NPV, the initial calculation suggested the need for 120 time units of observation.

However, to address the statistical challenge posed by the nonindependence of repeated measurements from individual participants—each contributing multiple observation periods—an intraclass correlation coefficient of 0.18 was used to adjust for within-participant correlation. The resulting design effect, reflecting this lack of independence, was calculated to be approximately 3. This design effect was used to multiply the basic sample size estimate, leading to a final requirement of 360 observational time periods. Consequently, a total cohort of 30 participants would satisfy this criterion, meeting the robust statistical power necessary for the intended analyses.

### Statistical Methods

Data analysis was conducted using SPSS (version 28; IBM Corp), alongside visualization in Tableau Desktop (Tableau Software) and data management in Excel 365 (Microsoft). Descriptive statistics were used to characterize the study data, using mean and SD for normally distributed variables and median and IQR for nonnormally distributed variables. The determination of normality was guided by kurtosis and skewness indices, with thresholds set at absolute values less than 1 for normal distribution.

Comparative analyses between groups were executed using the Student *t* test (2-tailed) for normally distributed data and the Mann-Whitney *U* test for data that deviated from normal distribution. The onset of an exacerbation was defined as the day on which a participant initiated an antibiotic treatment course; any subsequent antibiotic courses beginning within 10 days of the initial course's end were considered part of the ongoing exacerbation and not as discrete events.

The monitoring tool, a modified National Early Warning Score, was calculated with the exclusion of respiratory rate, as it could not be accurately measured remotely with the equipment provided to participants. Data were assessed longitudinally to establish individualized normal ranges for each participant by calculating variable SDs from the participant's average, while omitting any data from a 20-day span encompassing each exacerbation (10 days before and 10 days after).

An abnormal value occurring within the 10-day presymptomatic window preceding an exacerbation was classified as a true positive. Conversely, abnormal values within 10 days following an exacerbation were designated as "late" positives. Abnormal values outside of these windows were flagged as false positives. For the negative results, days featuring no abnormal values were declared true negatives unless they fell within a 10-day period preceding an exacerbation; in such cases, they were categorized as false negatives. We defined "episode sensitivity" for each variable as the detection of at least 1 abnormal result outside of the participant's normal range within the pre-exacerbation period.

### Ethical Considerations

Ethical approval was given by the NHS South Central Research Ethics Committee (14/SC/0298), and all participants gave written, informed consent. Participants were allowed to keep study equipment at the completion of the study but there was no financial compensation. Data were anonymized for analysis to safeguard participant information. Anonymized data can be made available for suitable studies on written request.

## Results

### Overview

A total of 30 participants were recruited, with study population baseline data shown in Table 1. A CONSORT (Consolidated Standards of Reporting Trials) diagram of study recruitment is shown in Figure S1 in Multimedia Appendix 1. A total of 39,534 physiological and 25,334 symptom data points were collected across 5358 participant-days. A total of 78 exacerbations were reported during the 6-month study period by 27 (90%) of 30 participants, with the median exacerbation count being 3 (IQR 1-3.75), and the remaining 3 (10%) participants not having exacerbations during the study period. During the study, 4 participants recorded a total of 6 hospital admissions for respiratory symptoms, giving an overall annualized rate of 0.2 per year compared to 0.6 per year (17 emergency hospital attendances from 7 participants) in the year prior to admission. Participant-level physiological and symptom data showed high interindividual differences around the point of exacerbation, as shown in Figures S2 and S3 in Multimedia Appendix 1.

Physiological data (Table 2) were compared using individualized *z* scores to adjust for interparticipant differences in exacerbation frequency. Values from exacerbation and 10 days on either side were compared with nonexacerbation days outside of this window. Peak flow rates and oxygen saturations were significantly lower during the exacerbation window, whereas weight and modified National Early Warning Score were significantly higher (all $P<.001$). Mean values of temperature, pulse rate, peak expiratory flow rate, and systolic blood pressure around exacerbation are shown in Figure 1.

Average appetite, breathing, cough, energy, and wellness symptom scores in the first week of data collection were correlated with overall SGRQ, giving correlation coefficients of –0.694, –0.761, –0.718, –0.798, and –0.805, respectively (all $P<.001$), showing good correlation between patient-reported symptom scores and a validated quality-of-life questionnaire. Symptom scores were also compared using individualized *z* scores to adjust for interparticipant differences in exacerbation frequency. Values from exacerbation and 10 days on either side

were compared with nonexacerbation days outside of this window. All symptom scores including the total symptom score were significantly lower at and around exacerbation compared to outside this range, as shown in Table 3 ($P$<.001 for each).

The trend of symptom scores around the point of exacerbation are shown in Figure 2, showing a gradual worsening in mean symptom score each day leading to the point of exacerbation, with gradual recovery following this. However, the magnitude of the decline is small, with each score decreasing by a mean of less than 1 point and by a similar magnitude.

The trends of total symptom scores through the study were examined. While the average total symptom score showed a slight improvement over time, this was not true for those with sputum *P aeruginosa*, who started with poorer symptom scores and showed a trend toward worsening over time, as shown in Figure 3.

**Table 1.** Study population baseline data. All participants with chronic obstructive lung disease (COPD) had a GOLD (Global Initiative for COPD) grade of D.

| Variable | Value (N=30) |
|---|---|
| Participants, n (%) | 30 (100) |
| Age (years), median (IQR) | 68.3 (61.3-73.6) |
| Female gender, n (%) | 17 (57) |
| **Smoking status, n (%)** | |
| Never smoked | 12 (40) |
| Ex-smoker | 17 (57) |
| Current smoker | 1 (3) |
| Pack-year history, median (IQR) | 15.5 (0-30) |
| Antibiotic courses per participant in the past 12 months, median (IQR) | 4 (3-5) |
| BMI (kg/m$^2$), mean (SD) | 26.3 (5.6) |
| Bronchiectasis, n (%) | 17 (57) |
| COPD, n (%) | 4 (13) |
| Bronchiectasis and COPD, n (%) | 9 (30) |
| Ischemic heart disease, n (%) | 4 (13) |
| Heart failure, n (%) | 4 (13) |
| Type 2 diabetes, n (%) | 2 (7) |
| FEV1[a] (% predicted), mean (SD) | 66.1 (28.4) |
| FVC[b] (% predicted), mean (SD) | 85 (23.4) |
| FEV1/FVC ratio, mean (SD) | 62.6 (16.0) |
| Sputum culture *Pseudomonas*, n (%) | 20 (67) |

[a]FEV1: forced expiratory volume in 1 second.

[b]FVC: forced vital capacity.

**Table 2.** Average results for physiological data throughout the study. *P* values were calculated from 2-tailed *t* test of the exacerbation period versus the nonexacerbation period for normally distributed data or Mann-Whitney U test for nonnormally distributed data. Exacerbation data were based on 27 participants, as 3 participants did not have an exacerbation.

| Physiological data | All data (n=30) | Exacerbation period (–10 to +10 days; n=27) | Normal range excluding the exacerbation period (n=27) | *P* value |
|---|---|---|---|---|
| Weight (kg), mean (SD) | 72.7 (13.9) | 74.5 (14.2) | 72.2 (14.2) | <.001 |
| Step count (per day), median (IQR) | 3065 (1242-7088) | 3296 (1285-6428) | 2975 (1183-7469) | .48 |
| Peak flow (L/min), median (IQR) | 233 (149-320) | 209 (133-289) | 243 (159-327) | <.001 |
| Oxygen saturation (%), median (IQR) | 95.4 (92.0-97.2) | 95.2 (92.0-97.0) | 95.5 (92.5-97.3) | <.001 |
| Systolic BP[a] (mm Hg), mean (SD) | 130 (22) | 130 (24) | 130 (22) | .90 |
| Diastolic BP (mm Hg), mean (SD) | 77 (12) | 77 (13) | 77 (12) | .88 |
| Pulse rate (per min), mean (SD) | 78.4 (13.2) | 78.8 (14.1) | 78.2 (11.9) | .052 |
| Temperature ($^{o}$C), median (IQR) | 36.7 (36.5-37.0) | 36.7 (36.5-37.0) | 36.6 (36.5-37.0) | .008 |
| mNEWS[b], mean (SD) | 1.4 (1.6) | 1.5 (1.7) | 1.3 (1.6) | <.001 |

[a]BP: blood pressure.

[b]mNEWS: modified National Early Warning Score.

**Figure 1.** Average (A) temperature, (B) pulse rate, (C) peak expiratory flow rate (PEFR), and (D) systolic blood pressure over 10 days on either side of the start of an exacerbation for all 78 exacerbations. bpm: beats per minute.

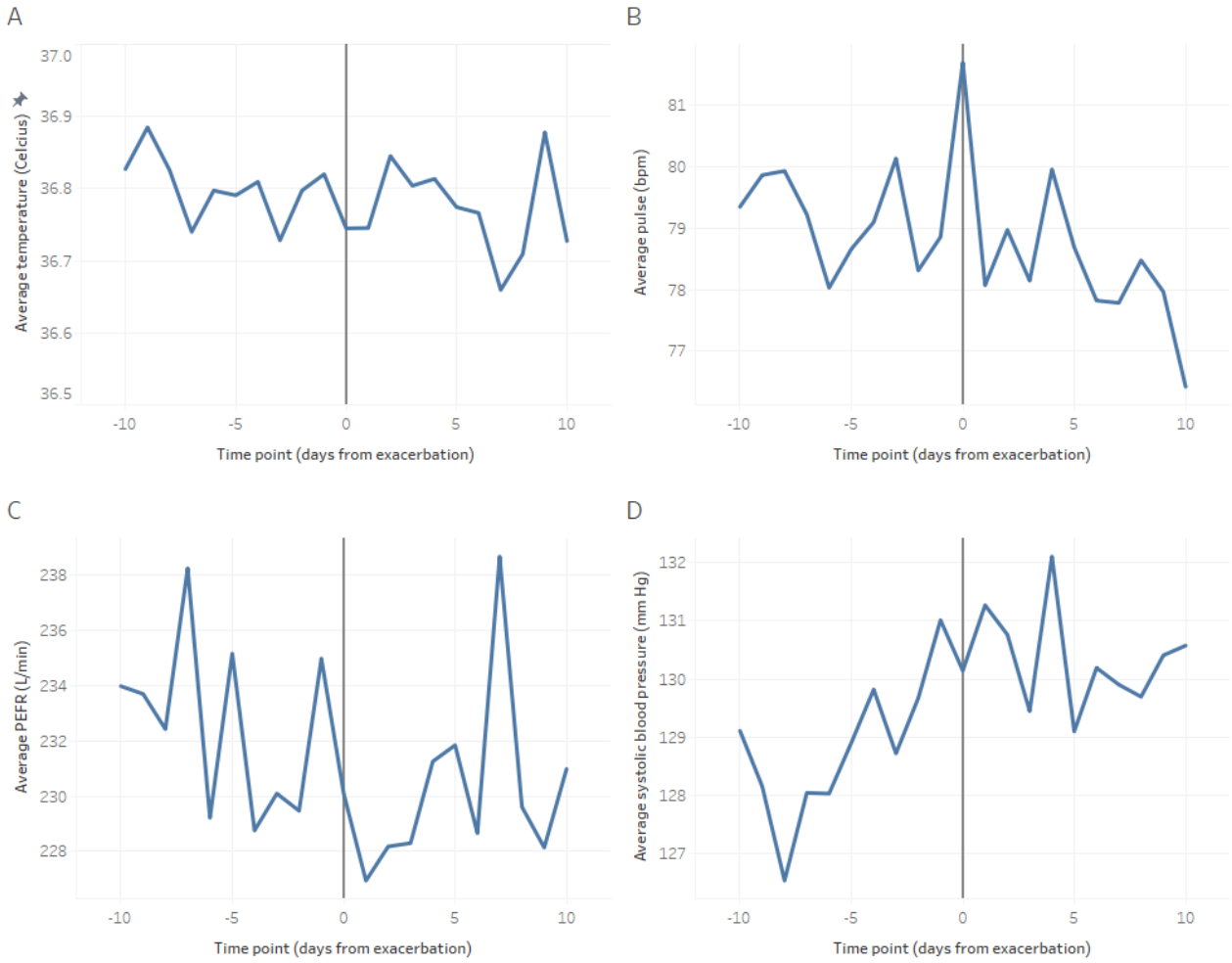
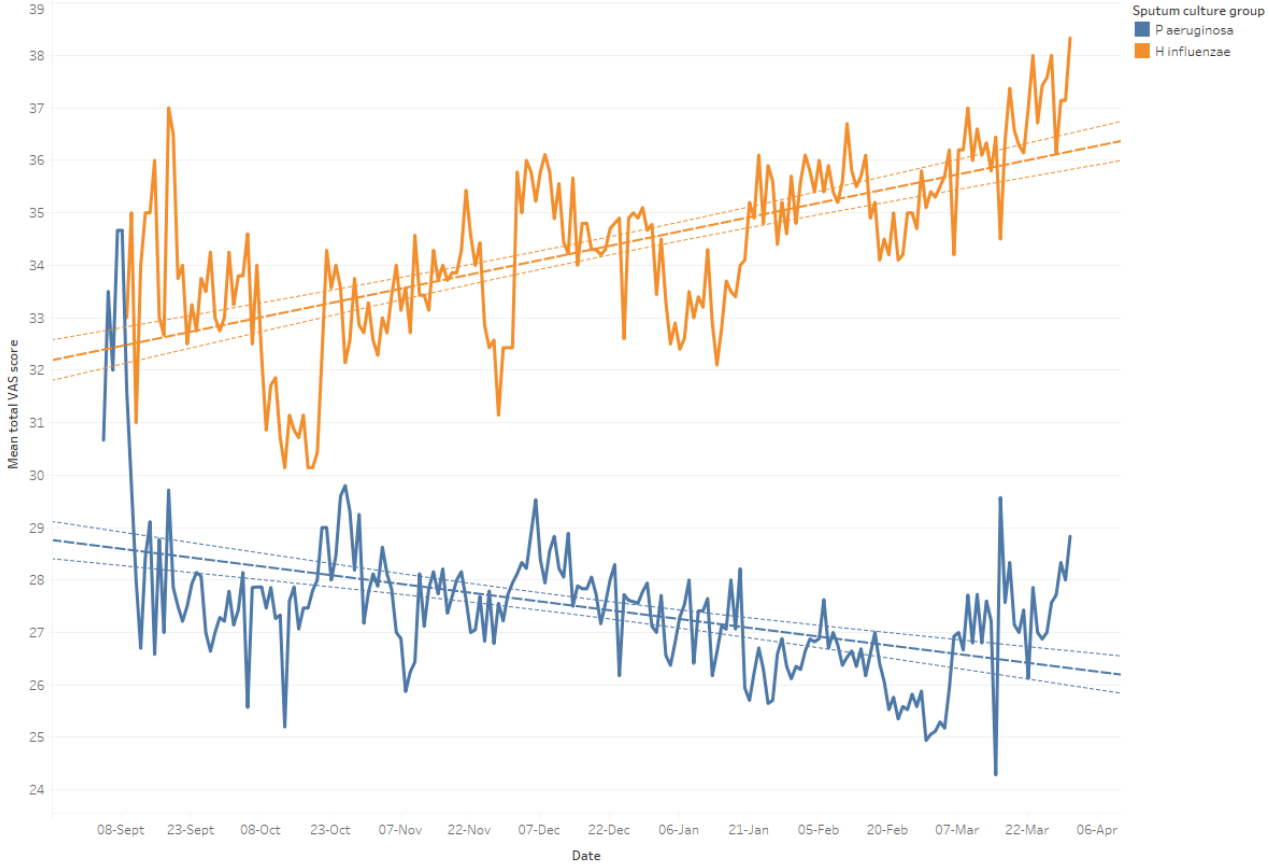


**Table 3.** Average results for all symptom data throughout the study and excluding exacerbations. *P* values were calculated from 2-tailed *t* test of the exacerbation period versus the nonexacerbation period.

| Symptom scores | All data (n=30) | Exacerbation period (–10 to 10 days; n=27) | Normal range excluding the exacerbation period (n=27) | *P* value |
|---|---|---|---|---|
| Wellness score, mean (SD) | 6.1 (1.8) | 5.7 (18) | 6.4 (1.8) | <.001 |
| Cough score, mean (SD) | 6.1 (1.8) | 5.6 (1.7) | 6.4 (1.8) | <.001 |
| Breathing score, mean (SD) | 6.0 (2.0) | 5.5 (2.0) | 6.2 (1.9) | <.001 |
| Appetite score, mean (SD) | 6.2 (2.0) | 5.8 (2.0) | 6.4 (2.0) | <.001 |
| Energy score, mean (SD) | 5.8 (2.1) | 5.2 (2.0) | 6.0 (2.1) | <.001 |
| Total symptom score, mean (SD) | 30.1 (9.2) | 27.7 (8.9) | 31.4 (9.1) | <.001 |

**Figure 2.** Mean symptom scores around the point of exacerbation (day 0) for all exacerbations.

**Figure 3.** Change in mean total symptom score by sputum culture, demonstrating worsening symptoms over time in the Pseudomonas aeruginosa group (lower line, n=20) and improving symptoms in the Haemophilus influenzae group (upper line, n=10). VAS: visual analog scale.




## Predictive Values

Individual control limits of variable widths were calculated as above. Table 4 shows predictive values at 2 SDs, which gave a balance between PPV and sensitivity. A range of control interval widths were calculated, and PPV for physiological variables were highest at 27.9% (at 3.3 SDs), while PPV for symptom variables peaked at 62.4% (at 4 SDs), which predicted 17.6% and 26.1% of episodes, respectively (Figures S4 and S5 in Multimedia Appendix 1). Individual-level data demonstrating the range of predictive values at control widths of 2 SDs for each participant are shown in Table S2 in Multimedia Appendix 1. Comparison between results for participants with COPD and bronchiectasis is shown in Table S3 in Multimedia Appendix 1, showing slightly higher PPV and episode sensitivity for participants with COPD but still less than 50%.

Physiological and symptom variables were analyzed using receiver operating characteristic analysis (Figure S6 in Multimedia Appendix 1), and no area under the curve (AUC) values were outside the range of 0.4-0.6, indicating no significant predictive value. This was repeated with *z* scores (Figures S7 and S8 in Multimedia Appendix 1) to normalize for individual variation, and all AUC values were within the 0.45-0.55 range, again showing no significant predictive value (Figure S9 in Multimedia Appendix 1). Receiver operating characteristic analysis was also conducted for COPD and bronchiectasis alone and for *P aeruginosa* and *H influenzae* colonization alone; in each analysis, no curve exceeded an AUC of 0.4-0.6, with the exception of peak flow in participants with *H influenzae*, which had an AUC of 0.374.

**Table 4.** Predictive values of abnormal results for physiological and symptom results with individualized control limits of 2 SDs.

| Physiological and symptom variables | PPV[a] (%) | NPV[b] (%) | Sensitivity (%) | Specificity (%) | Accuracy (%) | Episode sensitivity (%) |
|---|---|---|---|---|---|---|
| Weight | 27.23 | 82.76 | 7.26 | 95.82 | 80.14 | 24.36 |
| Steps | 13.51 | 82.51 | 3.46 | 95.36 | 79.45 | 19.23 |
| Peak flow rate | 26.18 | 83.06 | 6.90 | 95.91 | 80.46 | 28.21 |
| Oxygen saturation | 24.63 | 83.17 | 6.80 | 95.68 | 80.39 | 33.33 |
| Systolic BP[c] | 21.39 | 82.74 | 5.82 | 95.48 | 79.83 | 35.90 |
| Pulse rate | 20.00 | 82.94 | 5.44 | 95.48 | 79.99 | 37.18 |
| Temperature | 24.41 | 82.99 | 6.99 | 95.45 | 80.07 | 42.31 |
| Wellness score | 36.88 | 83.98 | 12.99 | 95.35 | 81.11 | 34.62 |
| Cough score | 36.81 | 84.28 | 16.06 | 94.22 | 80.69 | 42.31 |
| Breathing score | 37.19 | 84.51 | 18.07 | 93.61 | 80.52 | 43.59 |
| Appetite score | 40.29 | 84.27 | 14.99 | 95.35 | 81.44 | 47.44 |
| Energy score | 38.44 | 84.44 | 17.14 | 94.25 | 80.89 | 46.15 |
| Total symptom score | 38.46 | 84.28 | 15.39 | 94.85 | 81.11 | 43.59 |

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]BP: blood pressure.

## Discussion

Our study has revealed the limited predictive value of physiological markers and symptom scores in this population with airway disease and chronic colonization. The variability of physiological data and the subtlety of symptom changes challenge the reliance on these measures alone for predicting imminent exacerbations. Symptom scores on a population level indicated some deterioration around the time of exacerbation but were insufficient as stand-alone individual level predictors. Our approach to personalize reference ranges, accounting for intraparticipant variability, has been proven valuable but still fell short in enhancing predictive accuracy.

Despite these limitations, the role of these indicators cannot be entirely discounted. Physiological variables and symptom scores remain crucial components of a comprehensive disease monitoring plan for many respiratory diseases. The differentiation between physiological changes and symptom deterioration is important; while physiological parameters may not predict exacerbations with high accuracy, they provide valuable information on a patient's baseline health status, which can be crucial when responding to symptoms that suggest an exacerbation and therefore guide clinical decision-making, particularly when considered alongside an individual's typical symptomatology and exacerbation patterns. This approach was used in the PROMETE study for example, which showed that monitoring physiological values with respiratory physician reviews reduced hospital and emergency department attendances over 7 months, with more exacerbations being managed at home [24].

Moreover, the aggregated symptom score highlighted in our study, while not overtly predictive, could have potential applications when considering temporal trends over more extended periods, as some data suggest that telemonitoring is more beneficial over a longer term than in our study [25]. While acute predictive value is limited, monitoring this score could be useful for assessing overall disease management and quality of life across broader timeframes, which may be of interest in longitudinal studies. Systematic reviews of previous studies of the effectiveness of remote monitoring of airway disease have focused on COPD and have shown mixed results [26,27]. Of particular interest, physiological deterioration over time was noted in our study in those colonized by *P aeruginosa*, and this is consistent with previously reported data, suggesting a particularly high-risk group of patients [28].

Our study also reflects the multifaceted nature of exacerbations in chronic airway diseases, which likely result from a combination of factors, ranging from environmental triggers to individual patient behaviors. Furthermore, the findings underscore the essential role of education and self-management. The decrease in admission rates during the study, compared to the year prior, may indicate that equipping patients with self-management plans may help avert the most severe manifestations of exacerbations, thereby reducing hospital visits.

While our study has provided valuable insights into the management of chronic airway disease, the inherent limitations must be acknowledged. The relatively small cohort size and the specific inclusion of patients with known chronic airway colonization restrict the generalizability of the findings. The observational period did not extend across a full year, resulting in a potential underrepresentation of seasonal variations, although it crucially encompassed the winter months when exacerbations are generally more prevalent and severe. The exclusion of respiratory rate—a potentially significant physiological marker—from our data collection may have

omitted a critical variable with predictive capability. Additionally, while the symptom scores used in the study were less complex and showed good correlation with established measures such as the SGRQ, there remains the possibility that more nuanced symptom evaluation tools could yield stronger predictive correlations. Another notable aspect of the study was participant access to their own longitudinal data, which introduces the potential for reporting bias. However, the anticipated direction of such bias would more commonly lead to an overestimation of symptoms, resulting in false-positive identifications of exacerbations—a phenomenon not observed in our data. This suggests that while participant awareness of their data could be a confounding factor, its impact on the study's outcomes may be minimal. Despite these limitations, the study's strengths also warrant mention. By methodically tracking daily symptomatic and physiological changes within a clearly defined patient group, the study provides nuanced insight into the patterns preceding exacerbations. Moreover, the inclusion of the winter months offers pertinent data from a period of high clinical relevance due to the increased exacerbation risk.

Further studies examining respiratory rate, other symptom assessment, or continuous monitoring may be of benefit, but a strongly predictive single factor seems unlikely and alternative approaches to monitoring airway disease such as combining multiple sources of predictive data in the home setting are more likely to be useful.

## Acknowledgments

This work was supported by a grant from the Small Business Research Initiative through the East Anglia Health Science Network and Health Enterprise East.

## Authors' Contributions

All authors contributed to study design. Data analysis was performed by TLJ. The first draft of the manuscript was written by TLJ, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 751 KB - ijmr_v13i1e44397_app1.docx ]

## References

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2095-2128 [FREE Full text] [doi: 10.1016/S0140-6736(12)61728-0] [Medline: 23245604]
2. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
3. Spruit MA, Franssen FME, Rutten EP, Wopereis S, Wouters EF, Vanfleteren LE. A new perspective on COPD exacerbations: monitoring impact by measuring physical, psychological and social resilience. Eur Respir J 2016 Apr 01;47(4):1024-1027 [FREE Full text] [doi: 10.1183/13993003.01645-2015] [Medline: 27037307]
4. Halpin DMG, Decramer M, Celli B, Kesten S, Liu D, Tashkin DP. Exacerbation frequency and course of COPD. Int J Chron Obstruct Pulmon Dis 2012;7:653-661 [FREE Full text] [doi: 10.2147/COPD.S34186] [Medline: 23055714]
5. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 2002 Oct;57(10):847-852 [FREE Full text] [doi: 10.1136/thorax.57.10.847] [Medline: 12324669]
6. Amati F, Simonetta E, Gramegna A, Tarsia P, Contarini M, Blasi F, et al. The biology of pulmonary exacerbations in bronchiectasis. Eur Respir Rev 2019 Dec 31;28(154):190055 [FREE Full text] [doi: 10.1183/16000617.0055-2019] [Medline: 31748420]
7. Chalmers JD, Aliberti S, Filonenko A, Shteinberg M, Goeminne PC, Hill AT, et al. Characterization of the "frequent exacerbator phenotype" in bronchiectasis. Am J Respir Crit Care Med 2018 Jun 01;197(11):1410-1420. [doi: 10.1164/rccm.201711-2202OC] [Medline: 29357265]
8. Ställberg B, Lisspers K, Larsson K, Janson C, Müller M, Łuczko M, et al. Predicting hospitalization due to COPD exacerbations in Swedish primary care patients using machine learning - based on the ARCTIC Study. Int J Chron Obstruct Pulmon Dis 2021 Mar 16;16:677-688 [FREE Full text] [doi: 10.2147/COPD.S293099] [Medline: 33758504]

9. Davies G, Wells AU, Doffman S, Watanabe S, Wilson R. The effect of Pseudomonas aeruginosa on pulmonary function in patients with bronchiectasis. Eur Respir J 2006 Nov;28(5):974-979 [FREE Full text] [doi: 10.1183/09031936.06.00074605] [Medline: 16899482]
10. Parameswaran GI, Sethi S. Pseudomonas infection in chronic obstructive pulmonary disease. Future Microbiol 2012 Oct;7(10):1129-1132. [doi: 10.2217/fmb.12.88] [Medline: 23030418]
11. Wilkinson TMA, Donaldson GC, Hurst JR, Seemungal TAR, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004 Jun 15;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]
12. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation 2013 Apr;84(4):465-470 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.12.016] [Medline: 23295778]
13. Silcock DJ, Corfield AR, Gowens PA, Rooney KD. Validation of the National Early Warning Score in the prehospital setting. Resuscitation 2015 Apr;89:31-35 [FREE Full text] [doi: 10.1016/j.resuscitation.2014.12.029] [Medline: 25583148]
14. Warnier MJ, Rutten FH, de Boer A, Hoes AW, de Bruin ML. Resting heart rate is a risk factor for mortality in chronic obstructive pulmonary disease, but not for exacerbations or pneumonia. PLoS One 2014 Aug 26;9(8):e105152 [FREE Full text] [doi: 10.1371/journal.pone.0105152] [Medline: 25157876]
15. Walker PP, Pompilio PP, Zanaboni P, Bergmo TS, Prikk K, Malinovschi A, et al. Telemonitoring in chronic obstructive pulmonary disease (CHROMED). a randomized clinical trial. Am J Respir Crit Care Med 2018 Sep 01;198(5):620-628. [doi: 10.1164/rccm.201712-2404OC] [Medline: 29557669]
16. Janjua S, Carter D, Threapleton CJD, Prigmore S, Disler RT. Telehealth interventions: remote monitoring and consultations for people with chronic obstructive pulmonary disease (COPD). Cochrane Database Syst Rev 2021 Jul 20;7(7):CD013196 [FREE Full text] [doi: 10.1002/14651858.CD013196.pub2] [Medline: 34693988]
17. Adibi A, Sin D, Safari A, Johnson K, Aaron S, FitzGerald J, et al. The Acute COPD Exacerbation Prediction Tool (ACCEPT): a modelling study. Lancet Respir Med 2020 Oct;8(10):1013-1021. [doi: 10.1016/S2213-2600(19)30397-2] [Medline: 32178776]
18. Lee SD, Huang MS, Kang J, Lin CH, Park MJ, Oh YM, et al. The COPD Assessment Test (CAT) assists prediction of COPD exacerbations in high-risk patients. Respir Med 2014 Apr;108(4):600-608 [FREE Full text] [doi: 10.1016/j.rmed.2013.12.014] [Medline: 24456695]
19. Martinez-Garcia MA, Athanazio RA, Girón R, Máiz-Carro L, de la Rosa D, Olveira C, et al. Predicting high risk of exacerbations in bronchiectasis: the E-FACED score. Int J Chron Obstruct Pulmon Dis 2017 Jan 18;12:275-284 [FREE Full text] [doi: 10.2147/COPD.S121943] [Medline: 28182132]
20. Marin JM, Carrizo SJ, Casanova C, Martinez-Camblor P, Soriano JB, Agusti AGN, et al. Prediction of risk of COPD exacerbations by the BODE index. Respir Med 2009 Mar;103(3):373-378 [FREE Full text] [doi: 10.1016/j.rmed.2008.10.004] [Medline: 19013781]
21. Roberts C, Jones TL, Gunatilake S, Storrar W, Elliott S, Glaysher S, et al. The SENSOR study: protocol for a mixed-methods study of self-management checks to predict exacerbations of Pseudomonas aeruginosa in patients with long-term respiratory conditions. JMIR Res Protoc 2017 May 19;6(5):e89 [FREE Full text] [doi: 10.2196/resprot.6636] [Medline: 28526665]
22. Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease 2017 report. Global Initiative for Chronic Obstructive Lung Disease. 2017. URL: https://goldcopd.org/wp-content/uploads/2017/02/wms-GOLD-2017-FINAL.pdf [accessed 2024-08-05]
23. Seemungal TAR, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1998 May;157(5 Pt 1):1418-1422 [FREE Full text] [doi: 10.1164/ajrccm.157.5.9709032] [Medline: 9603117]
24. Segrelles Calvo G, Gómez-Suárez C, Soriano JB, Zamora E, Gónzalez-Gamarra A, González-Béjar M, et al. A home telehealth program for patients with severe COPD: the PROMETE study. Respir Med 2014 Mar;108(3):453-462 [FREE Full text] [doi: 10.1016/j.rmed.2013.12.003] [Medline: 24433744]
25. Lu JW, Wang Y, Sun Y, Zhang Q, Yan LM, Wang YX, et al. Effectiveness of telemonitoring for reducing exacerbation occurrence in COPD patients with past exacerbation history: a systematic review and meta-analysis. Front Med (Lausanne) 2021 Sep 10;8:720019 [FREE Full text] [doi: 10.3389/fmed.2021.720019] [Medline: 34568376]
26. Kruse C, Pesek B, Anderson M, Brennan K, Comfort H. Telemonitoring to manage chronic obstructive pulmonary disease: systematic literature review. JMIR Med Inform 2019 Mar 20;7(1):e11496 [FREE Full text] [doi: 10.2196/11496] [Medline: 30892276]
27. Li X, Xie Y, Zhao H, Zhang H, Yu X, Li J. Telemonitoring interventions in COPD patients: overview of systematic reviews. Biomed Res Int 2020 Jan 17;2020:5040521 [FREE Full text] [doi: 10.1155/2020/5040521] [Medline: 32016115]
28. Martínez-García MA, Soler-Cataluña JJ, Perpiñá-Tordera M, Román-Sánchez P, Soriano J. Factors associated with lung function decline in adult patients with stable non-cystic fibrosis bronchiectasis. Chest 2007 Nov;132(5):1565-1572. [doi: 10.1378/chest.07-0490] [Medline: 17998359]

## Abbreviations

**AUC:** area under the curve
**CONSORT:** Consolidated Standards of Reporting Trials
**COPD:** chronic obstructive pulmonary disease
**NPV:** negative predictive value
**PPV:** positive predictive value
**SGRQ:** St. George's Respiratory Questionnaire

*Edited by T de Azevedo Cardoso; submitted 17.11.22; peer-reviewed by L Zhou, N Maglaveras, J Edwards; comments to author 09.06.23; revised version received 02.09.23; accepted 20.05.24; published 08.10.24.*

*Please cite as:*
*Jones TL, Roberts C, Elliott S, Glaysher S, Green B, Shute JK, Chauhan AJ*
*Predictive Value of Physiological Values and Symptom Scores for Exacerbations in Bronchiectasis and Chronic Obstructive Pulmonary Disease With Frequent Exacerbations: Longitudinal Observational Cohort Study*
*Interact J Med Res 2024;13:e44397*
*URL: https://www.i-jmr.org/2024/1/e44397*
*doi:10.2196/44397*
*PMID:39378078*

Original Paper

# Temperature Measurement Timings and the Fever Detection Rate After Gastrointestinal Surgery: Retrospective Cross-Sectional Study

Shiqi Wang[1*], MD, PhD; Gang Ji[1*], MD, PhD; Xiangying Feng[1*], MD, PhD; Luguang Huang[2], MD, PhD; Jialin Luo[1], BMed; Pengfei Yu[1], MD, PhD; Jiyang Zheng[1], BMed; Bin Yang[1], MD, PhD; Xiangjie Wang[1], MD, PhD; Qingchuan Zhao[1*], MD, PhD

[1]Xijing Hospital of Digestive Diseases, Xijing Hospital, Fourth Military Medical University, Xi'an, China

[2]Medical Information Department, Xijing Hospital, Fourth Military Medical University, Xi'an, China

[*]these authors contributed equally

**Corresponding Author:**
Shiqi Wang, MD, PhD
Xijing Hospital of Digestive Diseases
Xijing Hospital
Fourth Military Medical University
127 Changlexi Street
Xi'an, 710032
China
Phone: 86 2984771533
Email: wsqfmmu@126.com

## Abstract

**Background:** Postoperative fever frequently indicates surgical complications and is commonly used to evaluate the efficacy of interventions against surgical stress. However, the presence of circadian rhythms in body temperature may compromise the accurate detection of fever.

**Objective:** This study aimed to investigate the detection rate of fever under intermittent measurement.

**Methods:** We retrospectively reviewed the clinical records of patients who underwent nonemergency gastrointestinal surgery between November 2020 and April 2021. Patients' temperature data were continuously collected every 4 seconds using a wireless axillary thermometer, and fever was defined as a temperature exceeding 38 °C within a day. To simulate intermittent measurement in clinical practice, the body temperature at each hour was selected from the continuously collected temperature dataset. Considering that temperatures are measured multiple times per day, all possible measurement plans using intermittent measurement were composed by combining 1-24 time points from the 24-hour daily cycle. Fever was clinically diagnosed based on the temperature readings at the selected time points per day. The fever detection rates for each plan, with varying measurement times, were listed and ranked.

**Results:** Based on the temperature data continuously collected by the thermometer, fever occurred in 60 (40.8%) of the 147 included patients within 3 days after surgery. Of the measurement plans that included 1-24 measurements daily, the fever detection rates ranged from 3.3% (2/60) to 85% (51/60). The highest detection rates and corresponding timings for measurement plans with 1, 2, 3, and 4 measurements daily were 38.3% (23/60; at 8 PM), 56.7% (34/60; at 3 AM and 7 or 8 PM), 65% (39/60; at 3 AM, 8 PM, and 10 or 11 PM), and 70% (42/60; at 12 AM, 3 AM, 8 PM, and 11 PM), respectively; and the lowest detection rates were 3.3% (2/60), 6.7% (4/60), 6.7% (4/60), and 8.3% (5/60), respectively. Although fever within 3 days after surgery was not correlated with an increased incidence of postoperative complications (5/60, 8.3% vs 6/87, 6.9%; $P$=.76), it was correlated with a longer hospital stay (median 7, IQR 6-9 days vs median 6, IQR 5-7 days; $P$<.001).

**Conclusions:** The fever detection rate of the intermittent approach is determined by the timing and frequency of measurement. Measuring at randomly selected time points can miss many fever events after gastrointestinal surgery. However, we can improve the fever detection rate by optimizing the timing and frequency of measurement.

***(Interact J Med Res 2024;13:e50585)*** doi:10.2196/50585

**KEYWORDS**

fever; gastrointestinal surgery; temperature measurement; temperature; detection; gastrointestinal; cross-sectional study

## Introduction

Fever commonly ensues following a diverse array of surgical interventions, including gastrointestinal procedures. Postoperative fever frequently stems from surgical stress and complicating factors [1,2]. In the context of gastrointestinal surgery, fever not only reflects elevated levels of surgical stress but may also signal the potential for complications, such as thrombosis, gastrointestinal leaks, intra-abdominal infections, and pulmonary infections [3-7]. In addition, fever manifestations are integral to the evaluation of targets and outcome measures for various Enhanced Recovery After Surgery (ERAS) interventions [8-14]. Therefore, precise fever detection is imperative for ensuring clinical safety and for the accurate appraisal of treatment efficacy. Time and accurate detection of fever promotes the early identification of patients at heightened risk for complications and supports the precise assessment of the efficacy of perioperative stress mitigation strategies.

To detect postoperative fever, patient temperatures are routinely measured at intervals of several hours against a predetermined threshold. A majority of health care facilities use a body temperature over 38 °C as the criterion for fever diagnosis [15]. This threshold is also commonly used in many clinical studies [5,16,17]. Nevertheless, there is often a failure to account for the timing of temperature measurements, which is pivotal since body temperature naturally oscillates, reaching a nadir at 6 AM and peaking between 4 PM and 6 PM [18]. As such, disparate measurement timings can result in substantial variance in the detected rates of fever within study cohorts. The question arises as to the optimal timing for temperature checks to identify fever. To date, neither a consensus nor guidelines exist to address this issue. Furthermore, the timing of temperature measurement is rarely specified in the literature on postoperative fever, with common practices including twice daily [17], during morning rounds [16], every 8 hours [19,20], or not report at all in certain studies [5,9,17,21-25]. This inconsistency in monitoring may overlook many febrile episodes, impeding the ability to evaluate the relationship between fever and surgical outcomes, as well as the effectiveness of perioperative antistress interventions.

Enhancing the frequency of temperature measurements could potentially improve fever detection rates; however, the practicality of such a method is low, incurring a substantial increase in medical workload for relatively little gain. Research has explored the use of wireless sensors for continuous real-time monitoring of patients' body temperatures. These sensors, transmitting temperature data to a central processor at frequent intervals, offer a more precise reflection of temperature fluctuations [26,27], which can facilitate prompt identification of postoperative complications [28]. Nonetheless, the functionality of these sensors is heavily dependent on reliable local network infrastructure and related devices. In settings bereft of sensor technology, alternative strategies must be considered. Harding et al [29] noted that fever patterns correspond to the diurnal variation in body temperature, peaking and troughing at consistent times, with nighttime fever incidence in the emergency department exceeding morning rates by a factor of 2.5. Such findings imply that adjusting measurement timings could improve fever detection rates. However, the optimal intervals and frequency of temperature assessments for maximal fever detection efficacy are yet to be determined. This study gathered data on the hour from a continuous temperature dataset and systematically constructed various hypothetical measurement schedules, subsequently evaluating the fever detection rate of each regimen to ascertain the most effective timing for temperature checks.

## Methods

### Study Design and Participants

This was a retrospective cross-sectional study. Between November 29, 2020, and April 1, 2021, consecutive patients who were aged 18 years or older and underwent nonemergency gastrointestinal surgery were included. Patients who took immunosuppressive drugs within 4 weeks before surgery, who had used antibiotics or antipyretic analgesics in the week before admission, and who were pregnant were excluded. To avoid the interference of nonsurgical fever, we also excluded patients who presented with fever before surgery. Patients whose temperature data were missing for any reason were also excluded. Demographic characteristics, surgical types, length of hospital stay, and in-hospital complications were collected.

### Ethical Considerations

This study was approved by the Ethics Committee of the First Affiliated Hospital of Air Force Medical University (approved KY20222271-C-1). The patients' data have been anonymized. As a retrospective study, there was no compensation, and the requirement for informed consent was waived.

### Body Temperature Measurement

Body temperature was continuously measured every 4 seconds by a wireless axillary thermometer (iThermonitor; Raiing Medical Company). The measurement accuracy of the sensor is 0.01 °C, and the readings are consistent with those of mercury thermometers [30]. On admission, a hypoallergenic adhesive patch (Raiing Medical Company) was used to securely position the iThermonitor in the shaved axilla of the patient. The temperature data were transmitted to a repeater through low-energy Bluetooth and then transferred to a central workstation, where the data were saved on an electronic monitoring panel. A detailed description of the technical parameters of the iThermonitor can be found in Multimedia Appendix 1. The temperature data on the day of surgery were dismissed, and the temperature data of the next 3 days were retrieved and named the first-, second-, and third-day temperatures.

### End Points

Taking into consideration previous research reports and consensus outcomes, fever was defined as a body temperature that exceeded 38 ˚C [5,15-17]. Based on the definition of fever

and the data conscientiously collected by the sensors, the fever incidence on each day of the first 3 days and the total fever incidence of the first 3 days were investigated.

Patients were divided into a fever group and a nonfever group, and the clinical outcomes were compared between the 2 groups. The correlation between fever on the first day and fever in the next 2 days was investigated.

### Intermittent Measurement Simulation

In the simulated clinical temperature measurement analysis, we included all patients who were determined to have a fever based on sensor temperature data. In clinical practice, for ease of implementation and documentation, temperature measurements are typically taken on the hour. Therefore, the temperature data for every hour were selected from the consecutively collected dataset.

A brute force strategy was used to list the fever detection rate of every possible measurement plan with varied measurement time points per day. The simulated intermittent measurement plans, where every possible temperature measurement plan in clinical practice, including 1 to 24 time points, were composed of n time points from the 24 hours, such as C (24, 1), C (24, 2), C (24, 3), C (24, 4), C (24, 6), C (24, 8), C (24, 12), and C (24, 24). If the temperature data at any time points included in the temperature measurement plan exceed 38 °C, it was considered that the clinical temperature measurement plan has detected a fever.

For example, C (24, 2) means diagnosing fever based on temperatures at any 2 time points in 24 hours, such as 12 AM and 1 AM, 12 AM and 2 AM, or 12 AM and 3 AM. The total number of combinations of C (24, 2) was 276. Fever was clinically diagnosed based on the temperature at the selected time points, and the fever detection rates of every plan with varied measurement times were listed and ranked.

### Statistical Analysis

Temperature data processing, including the combination of the measurement timings, calculation of fever incidence, and calculation of fever detection rate, was managed by Python (version 3.7.3; Python Software Foundation), pandas (version 1.1.3; Pandas Development Team), and NumPy (version 1.21.6; NumPy Community). The statistical analysis was conducted by SPSS 22.0 (IBM Corp). Categorical data are reported as numbers with proportions, and quantitative data are reported as the mean with SD or, where appropriate, as the median with an IQR. Categorical data were compared using the chi-square test or Fisher exact test, where appropriate. For continuous data, the Student *t* test (2-tailed) or Mann-Whitney *U* test was used. A 2-sided *P* value of <.05 was considered statistically significant. Because of the exploratory nature of this survey, the sample size calculation was not performed.

## Results

### Patient Characteristics and Clinical Outcomes

A total of 147 patients who underwent gastrointestinal surgery were included. All the patients had complete temperature data within 3 days after surgery, and no missing values needed to be processed. Temperature data that were continuously collected by the sensor were used. Fever was detected in a total of 40.8% (60/147) patients within 3 days after surgery. Table 1 shows the demographic characteristics, surgical types, length of hospital stay after surgery, and in-hospital complications. The median age of the patients was 60 (IQR 53-67) years, and 62.6% (92/147) of the patients were male. The median length of hospital stay was 6 (IQR 5-8) days. Compared with the patients without fever within 3 days after surgery, patients with fever experienced a longer length of hospital stay (median 7, IQR 6-9 days vs median 6, IQR 5-7 days; *P*<.001). No significant difference in the postoperative complication rate was found between patients with and without fever within the 3 days (5/60, 8.3% vs 6/87, 6.9%; *P*=.76).

**Table 1.** Patient characteristics.

| Characteristics | All patients (N=147) | Fever (n=60) | No fever (n=87) | *P* value |
|---|---|---|---|---|
| **Age (years), median (IQR)** | 60 (53-67) | 60.5 (56-66) | 58 (50-69) | .30[a] |
| **Sex (male), n (%)** | 92 (62.6) | 36 (60) | 56 (64.4) | .59[b] |
| **Comorbidity, n (%)** | | | | |
| Hypertension | 28 (19) | 12 (20) | 14 (16.1) | .54[b] |
| Diabetes | 11 (7.5) | 4 (6.8) | 7 (8) | >.99[c] |
| Coronary artery disease | 12 (8.2) | 5 (8.3) | 7 (8) | >.99[c] |
| **Laparoscopic surgery, n (%)** | 81 (55.1) | 34 (56.7) | 47 (54) | .75[b] |
| **Surgery types, n (%)** | | | | .15[c] |
| Esophagectomy | 5 (3.4) | 2 (3.3) | 3 (3.4) | |
| Gastrectomy | 60 (40.8) | 32 (53.3) | 28 (32.2) | |
| Colorectal resection | 72 (49) | 24 (40) | 48 (55.2) | |
| Small intestinal resection | 7 (4.8) | 2 (3.3) | 5 (5.7) | |
| Pancreaticoduodenectomy | 2 (1.4) | 0 (0) | 5 (5.7) | |
| Pancreatectomy | 1 (0.7) | 0 (0) | 1 (1.1) | |
| **Length of hospital stay (days), median (IQR)** | 6 (5-8) | 7 (6-9) | 6 (5-7) | <.001[d] |
| **Complications, n (%)** | 11 (7.5) | 5 (8.3) | 6 (6.9) | .76[c] |
| Pneumonia | 4 (2.7) | 2 (3.3) | 2 (2.3) | |
| Intestinal obstruction | 4 (2.7) | 2 (3.3) | 2 (2.3) | |
| Leakage | 3 (2) | 1 (1.7) | 2 (2.3) | |
| Incision dehiscence | 2 (1.4) | 1 (1.7) | 1 (1.1) | |

[a]Student *t* test (2-tailed).

[b]Chi-square test.

[c]Fisher exact test.

[d]Mann-Whitney *U* test.

## Fluctuation of Body Temperature and Fever Detection Rate

The fluctuations in body temperature on the first, second, and third days after surgery are shown in Figure 1A-C. The mean body temperature ranged from 36.88 °C to 37.24 °C, and the mean body temperatures in the first 3 days after surgery were 37.02 °C, 37.08 °C, and 37.06 °C, respectively.

The average body temperature in 1 day is shown in Figure 1D**,** and the body temperature varied throughout the day, with its nadir at 8 AM and its zenith at 11 PM. Figure 2 shows the fever detection rate by each hour. Within 24 hours, 38.3% (23/60) fever was detected by taking measures at 7 PM and 8 PM, and only 3.3% (2/60) fever was detected by taking measures at 1 PM.

**Figure 1.** Body temperature curve after gastrointestinal surgery. Body temperature fluctuations on the (A) first, (B) second, and (C) third days after surgery. The average body temperatures in the first 3 days after surgery were 37.02 °C, 37.08 °C, and 37.06 °C, respectively. (D) Average body temperature within 24 hours. The average body temperature bottomed out at 8 AM and peaked at 11 PM.

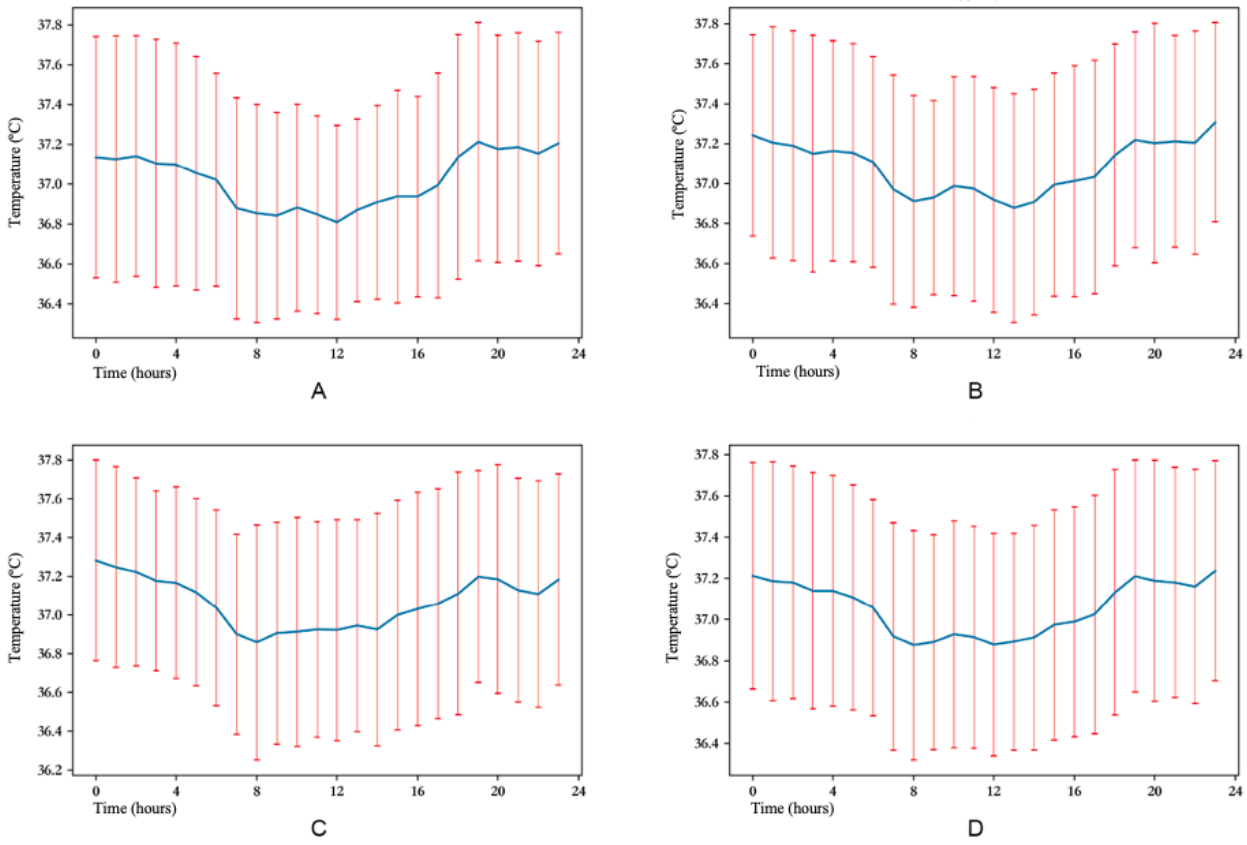
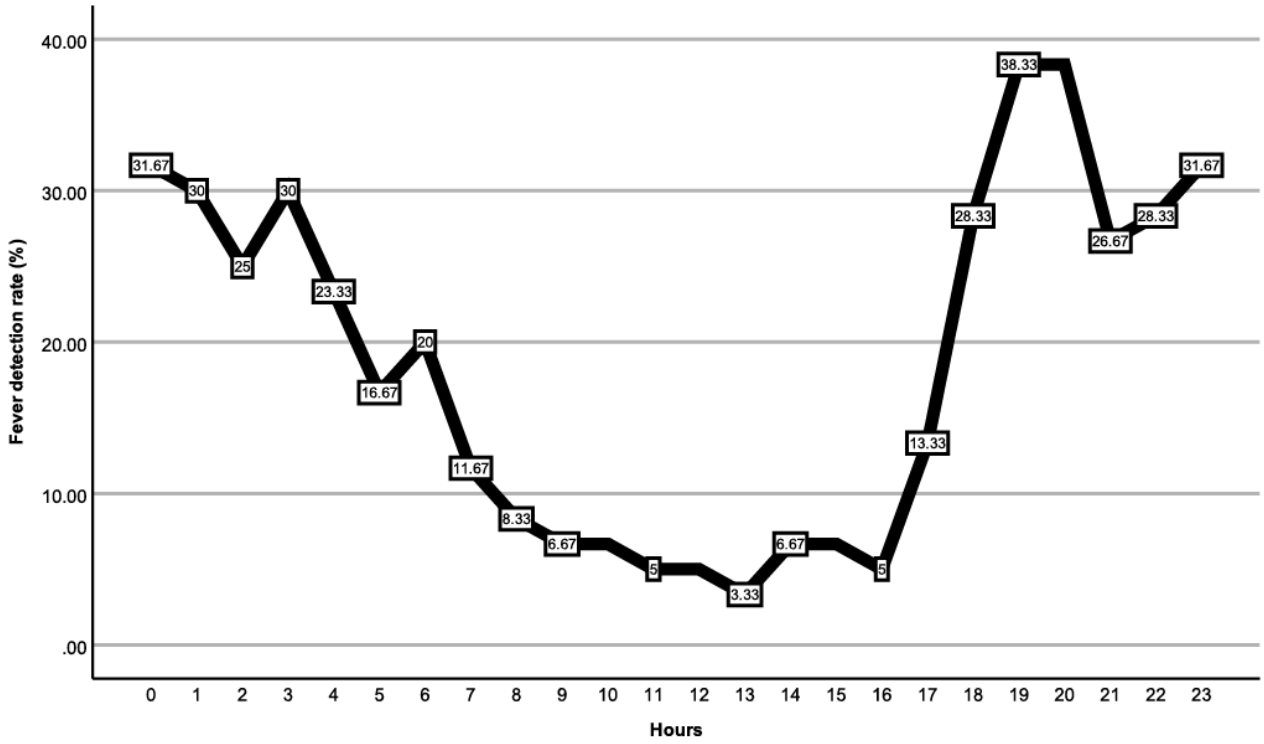



**Figure 2.** Fever detection rate on each hour. The vertical axis represents the fever detection rate, and the abscissa axis represents the hours. The fever detection rate peaked at 38.3% at 7 PM and 8 PM and reached a nadir of 3.3% at 1 PM.



## Fever Detection Rate and Measurement Times

Using intermittently collected temperature data on the hour, varied measurement plans were constructed and demonstrated. The highest and lowest detection rates and the measurement timings for the highest detection rate are shown in Table 2. For the 1–time point model C (24, 1), meaning fever was diagnosed by the temperature data at 1 time point per day, the fever detection rate ranged from 3.3% (2/60) to 38.3% (23/60). In the 2–time point model C (24, 2), meaning fever was diagnosed by the temperature data at 2 time points per day, the fever detection rate ranged from 6.7% (4/60) to 56.7% (34/60). For the C (24, 3), C (24, 4), and C (24, 6) models, the fever detection rate ranged from 6.7% (4/60) to 65% (39/60), from 8.3% (5/60) to 70% (42/60), and from 11.7% (7/60) to 76.7% (46/60), respectively. When the measurement frequency was increased to hourly, the detection rate gradually reached a plateau of 85% (51/60) (Figure 3A).

**Figure 3.** Correlation between fever detection rate and measurement times. The vertical axis represents the fever detection rate, and the abscissa axis represents the measurement times. The blue (D1), orange (D2), and green (D3) lines show the fever detection rates on the first, second, and third days after surgery, respectively. The red line (D1-3) shows the fever detection rate within 3 days. (A) The averaged fever detection rates with varied time points. (B) The highest detection rates with varied time points.

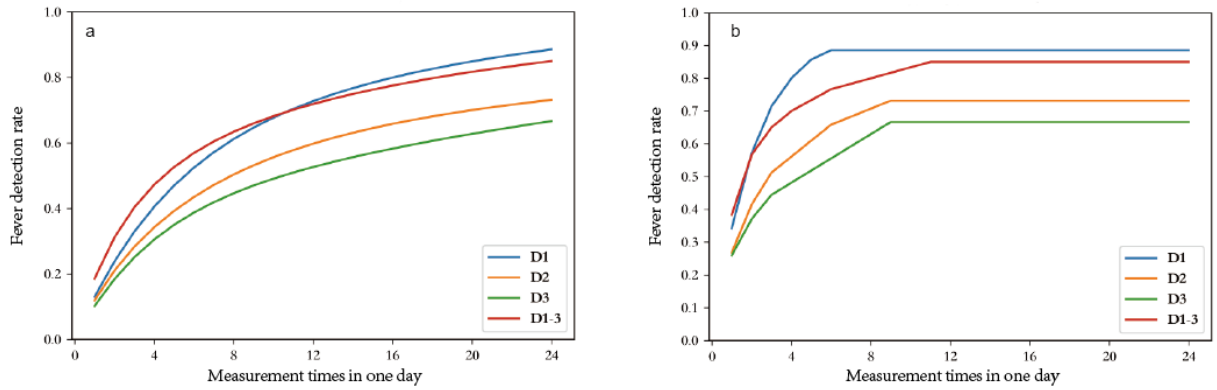



**Table 2.** Fever detection rates of body temperature measurement plans with varied measurement timings.

| Measurement timings | Fever detection rate[b] (%, 95% CI) | | Measurement plans with the highest detection rate |
|---|---|---|---|
| | Lowest | Highest | |
| **Measurement plans [a]** | | | |
| C (24, 1) | 3.3 (0.4-6.2) | 38.3 (30.5-46.2) | 7 or 8 PM |
| C (24, 2) | 6.7 (2.6-10.7) | 56.7 (48.7-64.7) | 3 AM and 7 or 8 PM |
| C (24, 3) | 6.7 (2.6-10.7) | 65.0 (57.3-72.7) | 3 AM, 8 PM, and 10 or11 PM |
| C (24, 4) | 8.3 (3.9-12.8) | 70.0 (62.6-77.4) | 12 AM, 3 AM, 8 PM, and 11 PM |
| C (24, 6) | 11.7 (6.5-12.9) | 76.7 (69.8, 83.5) | 12 AM, 3 AM, 6 AM, 4 PM, 8 PM, and 11 PM |
| C (24, 8) | 18.3 (12.1-24.6) | 80.0 (73.5-86.5) | 12 AM, 1 AM, 3 AM, 5 AM, 6 AM, 4 PM, 8 PM, 11 PM, etc (n=26) |
| C (24,24) | 85.0 (79.2-90.8) | 85.0 (79.2-90.8) | —[c] |
| **Plans in our ward** | | | |
| Plan A | 43.3 (40.3-56.4) | 43.3 (40.3-56.4) | 6 AM and 6 PM |
| Plan B | 48.3 (40.3-56.4) | 48.3 (40.3-56.4) | 6 AM, 10 AM, 2 PM, and 6 PM |
| Plan C | 58.3 (50.4-66.3) | 58.3 (50.4-66.3) | 6 AM, 10 AM, 2 PM, 6 PM, 10 PM, and 2 AM |

[a]C (24, r), selecting r time points from the 24 hours in 1 day.

[b]Fever detection rate = (detected number of patients with fever) / (all patients with fever) within 3 days after surgery.

[c]Not applicable.

## Fever Detection Rate and Measurement Timings

The detection rate of the intermittent approach is influenced by the measurement timings. Table 2 shows that the corresponding time points of the top detection rates were distributed throughout the nighttime. For example, at 7 or 8 PM for C (24, 1); 3 AM and 7 or 8 PM for C (24, 2); 3AM, 8 PM, and 10 or 11 PM for C (24, 3); and 12 AM, 3 AM, 8 PM, and 11 PM for C (24, 4), the detection rate also reached a plateau by taking measures at fewer specific time points (Figure 3B). For example, an 85% (51/60) detection rate can also be achieved by using 11 time

points: 12 AM, 1 AM, 3 AM, 5 AM, 6 AM, 8 AM, 4 PM, 5 PM, 8 PM, 9 PM, and 11 PM (see row C (24, 11) in Multimedia Appendix 2). However, this is also cumbersome in real clinical work.

In our ward, according to the nursing grade, postoperative temperature are measured with 3 plans: plan A (6 AM and 6 PM), plan B (6 AM, 10 AM, 2 PM, and 6 PM), and plan C (6 AM, 10 AM, 2 PM, 6 PM, 10 PM, and 2 AM). Based on the continuously collected data, the fever detection rate of these plans were 43.3% (26/60), 48.3% (29/60), and 58.3% (35/60), respectively. The optimal detection rate of plans with the same measurement times were 56.7% (34/60), 70% (42/60), and 76.7% (46/60), respectively (Table 2).

# Discussion

## Principal Findings

This is the first study to investigate fever detection rates by intermittent temperature measurements. In this study, every possible intermittent measurement plan with varied measurement timings was constructed, and the corresponding fever detection rates were calculated. The results showed that fever was less frequently detected by medical staff than by the sensors, and the upper limit of detection rates by intermittent measurement was 85% (51/60) when body temperature was measured every hour. For measurement plans with varied daily frequencies, we can improve the detection rates by adjusting the measurement timings.

## Limitations

While our findings have important implications, we acknowledge several limitations to our study. First, this is a small-sample retrospective study; as such, its results may be subject to bias. Second, we included patients who had undergone gastrointestinal surgery; it remains to be verified whether patients who have undergone other types of surgery also exhibit similar postoperative body temperature characteristics, which would require validation in other patient cohorts. Third, the detection rates were calculated based on the assumption that temperature was measured on the hour. There may be better time points at which fever detection is the highest. However, for convenience, body temperature is usually measured during the hour of clinical work, and our assumption is consistent with clinical practice. Fourth, previous studies and our research both indicate that the incidence of febrile events decreases as the duration of hospitalization increases [31-33]. Given that our study had a small sample size, and febrile events became infrequent after 3 days, it was challenging to discern the differences across various temperature monitoring schedules. Therefore, we chose to analyze the time period during which the occurrence of febrile events was higher, and only body temperature within the first 3 days after surgery was recorded. Whether the selected time points that were determined in this study are applicable after 3 days remains to be determined in further studies. Fifth, like most studies, we used a fixed threshold to define fever at different times. However, considering the variability of body temperature, it might be more reasonable to use a floating threshold to determine whether a patient is having a fever at different times. For instance, whether a body temperature exceeding 37.5 ˚C after waking up, or exceeding 37 ˚C, should also be considered an abnormal state. Nonetheless, we currently lack a more rational method to define fever. Moreover, this issue goes beyond the interpretive scope of this study and requires further exploration in future research. In addition, body temperature is influenced by age, sex, and even weather [34]. However, stratified analysis was not performed since it is not practical to do so in the ward to define fever by varied levels.

## Comparison with Previous Work

Consistent with a previous study which found that the proportion of patients with fever increased 2.4 to 3.6 times from morning to evening [29], we also observed that the timings of the measurement plans with the highest detection rates were predominantly at night. Moreover, as Figure 2 illustrates, the discrepancy between fever detection rates during the day and at night was more pronounced (2/60, 38.3% at 7 or 8 PM vs 2/60, 3.3% at 1 PM). This can be explained by the circadian rhythm of human body temperature. It is well known that the body temperature fluctuates throughout the day [18], with potential fluctuations of up to 1 °C within a single day [35]. In our study, we also found that body temperature exhibited rhythmic variations in patients who underwent gastrointestinal surgery. Figure 1D demonstrates that patients' body temperatures after surgery tend to be higher at night and lower during daylight, with the minimum recorded at 8 AM and the maximum at 11 PM. Hence, assessing fevers using temperature readings taken at various times throughout the day can lead to substantial discrepancies in conclusions.

Unfortunately, the timing of temperature monitoring is often neglected in both clinical research and practical settings. Notably, even in medical students' textbooks, there is no clear protocol for monitoring body temperature during the perioperative period. Medical centers tend to formulate postoperative temperature monitoring protocols based on customary local practices rather than standardized guidelines. This study demonstrates that our hospital's long-standing protocol has failed to effectively identify postoperative fever events. As indicated by Table 2, conducting as many as 6 temperature measurements daily only detected 58.3% (35/60) of patients with fever. If we adjust the temperature measurement times to 3 AM and 7 or 8 PM, 2 daily measurements could still identify 56.7% (34/60) of the cases. Optimizing the schedule to include checks at 12 AM, 3 AM, 6 AM, 4 PM, 8 PM, and 11 PM could improve fever detection rates to 76.7% (46/60).

As with our usual practice, some studies on the clinical significance of fever are typically measured only at a few unreported times of the day [16,17,19,20], while other studies do not show the timing of measurement [6,9,17,21-25]. Assuming body temperature is measured per hour, there would be 24 time points per day, resulting in a vast array of measurement schedules. Given the wide range of detection rates among the numerous measurement schedules, many patients with fever might go unidentified when measurements are taken at randomly selected times. Since fevers cannot be accurately detected, the interpretation of the clinical significance of postoperative fever may also be biased. In addition, we also see

some clinical studies that consider fever as an outcome of the intervention, especially those related to ERAS strategies for perioperative stress control [8-14]. In these studies, the timing and frequency of temperature measurements are not reported either. If consistency of the timing of temperature measurements is not considered when assessing fever, biases are likely to occur when evaluating the efficacy of the respective clinical interventions. Therefore, we may consider including recommendations on the timing of temperature measurements in ERAS-related guidelines.

Multimedia Appendix 2 lists the optimal measurement schedules that achieve the highest detection rates, varying from once to 23 times per day. By aligning these schedules with the routine practices of local hospitals, clinicians can formulate more precise thermometric protocols. For instance, a tridaily measurement regimen might entail taking temperatures at 3 AM, 8 PM, and 10 or 11 PM, as specified in row C (24, 3) in Table 2. It is essential to recognize, however, that these proposed times are flexible rather than absolute mandates for fever surveillance. In clinical practice, temperatures will also be measured at any necessary time. Given the symptoms that accompany fever, it is reasonable to measure temperature at suggested times as well as when needed.

What is the clinical significance of detecting postoperative fever, especially fever that occurs in the early days following surgery? Many studies have investigated the correlation between postoperative fever and infection. Most have found that postoperative fever is a marker of infection with very low specificity and sensitivity [2,5]. Some researchers have reported that among patients who developed fever after abdominal surgery, only 2% had positive blood cultures [5]. Among patients with fever following orthopedic surgeries, the positive finding rates for chest x-rays, urinalysis, urine cultures, and blood cultures were 0.3%, 28.5%, 10.9%, and 3.5%, respectively. Such low cost-effectiveness has led some researchers to question the use of postoperative temperature measurement [36]. One study even instructed the clinical team responsible for patient care to remain unaware of the patient's body temperature and required clinical decisions to be made without looking at temperature data. This study reported a positive predictive value of merely 8% for fever as an indicator of infection, suggesting the potential abandonment of routine temperature measurements [17]. Our findings align with these observations, demonstrating an insignificant link between fevers within the first 3 postoperative days and the onset of complications.

Although the prevalence of postoperative fever may not require immediate imaging or bacteriological assessments, it is inadvisable to ignore it and leave patients to manage the condition without support. The risks posed by postoperative fever extend beyond infection. Postoperative fever is also associated with the release of inflammatory mediators in the absence of infection. This study indicates that fever during the early postoperative days has a positive correlation with prolonged hospital stays (with a median of 7, IQR 6-9 days, compared with a median of 6, IQR 5-7 days). The longer hospitalization might be attributable to surgical stress, as a fever following surgery may arise from inflammation and tissue damage [37], suggesting that patients enduring pronounced surgical stress may need additional recovery time. Monitoring for postoperative fever is crucial in evaluating the magnitude of surgical stress and the efficacy of interventions to mitigate it. Fever, as a surgical stressor, constitutes a postoperative adverse event and an unpleasant experience that necessitates closer nursing attention. Considering the benign nature of early fever, there might be 2 approaches, refraining from intervention and allowing the fever to subside on its own, or providing necessary medical care, such as pain management [38], physical cooling, physical examinations, and psychological comfort, to facilitate the recovery process. If we opt to take some action, routine ward rounds could be considered during peak fever times, such as between 7 PM and 8 PM.

## Conclusions

In conclusion, reliance on traditional, arbitrary temperature measurement can lead to the oversight of numerous febrile episodes. From the standpoint of both clinical safety concerning fevers and the interpretability of clinical research, it is necessary to improve the detection rate of postoperative febrile events. Even in medical settings where continuous temperature monitoring sensors are unavailable, adjusting the timing for measuring temperatures to the nighttime can substantially improve the detection of febrile events. Postoperative body temperature monitoring protocols can be revised in accordance with the working habits of local hospitals. In addition, to facilitate more precise assessments of study outcomes, future research examining postoperative fevers should consider detailing the timing of temperature recordings in their reports.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant 82100680) and the Research Project Fund of Xijing Hospital (grant XJZT21CM40).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of iThermonitor wireless temperature monitoring technology.
[DOCX File , 706 KB - ijmr_v13i1e50585_app1.docx ]

Multimedia Appendix 2
Fever detection rates of measurement plans with varied measurement timings.
[DOCX File , 21 KB - ijmr_v13i1e50585_app2.docx ]

## References


1. Pile JC. Evaluating postoperative fever: a focused approach. Cleve Clin J Med 2006;73 Suppl 1:S62-S66. [doi: 10.3949/ccjm.73.suppl_1.s62] [Medline: 16570551]
2. Uçkay I, Agostinho A, Stern R, Bernard L, Hoffmeyer P, Wyssa B. Occurrence of fever in the first postoperative week does not help to diagnose infection in clean orthopaedic surgery. Int Orthop 2011;35(8):1257-1260 [FREE Full text] [doi: 10.1007/s00264-010-1128-z] [Medline: 20871993]
3. Endo M, Tanaka Y, Sato Y, Ohno S, Yoshida K. Asymptomatic pulmonary thromboembolism diagnosed based on prolonged fever after gastric cancer surgery: a case report with literature review. Int J Surg Case Rep 2022;92:106836 [FREE Full text] [doi: 10.1016/j.ijscr.2022.106836] [Medline: 35219115]
4. Winter JM, Cameron JL, Yeo CJ, Lillemoe KD, Campbell KA, Schulick RD. Duodenojejunostomy leaks after pancreaticoduodenectomy. J Gastrointest Surg 2008;12(2):263-269. [doi: 10.1007/s11605-007-0370-1] [Medline: 17968631]
5. Moreira ADL, Vogel JD, Kalady MF, Hammel J, Fazio VW. Fever evaluations after colorectal surgery: identification of risk factors that increase yield and decrease cost. Dis Colon Rectum 2008;51(5):508-513. [doi: 10.1007/s10350-007-9183-2] [Medline: 18228099]
6. Çetin DA, Gündeş E, Çiyiltepe H, Aday U, Uzun O, Cumhur Değer K, et al. Risk factors and laboratory markers used to predict leakage in esophagojejunal anastomotic leakage after total gastrectomy. Turk J Surg 2019;35(1):6-12 [FREE Full text] [doi: 10.5578/turkjsurg.4117] [Medline: 32550297]
7. Kim JH, Kim J, Lee WJ, Seong H, Choi H, Ahn J, et al. The incidence and risk factors for surgical site infection in older adults after gastric cancer surgery: a strobe-compliant retrospective study. Medicine (Baltimore) 2019;98(32):e16739 [FREE Full text] [doi: 10.1097/MD.0000000000016739] [Medline: 31393386]
8. Chen Y, Yang J. Perioperative fast-track surgery nursing intervention for patients with kidney stone disease under computed tomography imaging. Contrast Media Mol Imaging 2023;2023:1101388 [FREE Full text] [doi: 10.1155/2023/1101388] [Medline: 36793497]
9. Booth A, Ford W, Olajire-Aro T, George V, Curran T. Fever after colectomy in the enhanced recovery era: incidence and risk factors for pathologic source. J Surg Res 2022;277:171-180. [doi: 10.1016/j.jss.2022.04.024] [Medline: 35500512]
10. Park SH, Kang SH, Lee SJ, Won Y, Park YS, Ahn SH, et al. Actual compliance rate of enhanced recovery after surgery protocol in laparoscopic distal gastrectomy. J Minim Invasive Surg 2021;24(4):184-190 [FREE Full text] [doi: 10.7602/jmis.2021.24.4.184] [Medline: 35602855]
11. Abet E, Drissi F, Couëtte C, Jean MH, Denimal F, Podevin J, et al. Predictive value of inflammatory markers for postoperative recovery following colorectal surgery. Int J Colorectal Dis 2020;35(6):1125-1131. [doi: 10.1007/s00384-020-03594-y] [Medline: 32291509]
12. Li Q, Wang J, Zhang G, Wang J, Yang B, Zhang Z. Feasibility and safety comparison of laparoscopy-assisted versus open gastrectomy for advanced gastric carcinoma with D2 lymphadenectomy. Jpn J Clin Oncol 2016;46(4):323-328 [FREE Full text] [doi: 10.1093/jjco/hyw001] [Medline: 26851297]
13. Ding J, Liao G, Xia Y, Zhang ZM, Pan Y, Liu S, et al. The necessity of indwelling gastrointestinal decompression after gastrectomy: a meta-analysis. J Surg Res 2013;179(1):e71-e81. [doi: 10.1016/j.jss.2012.02.030] [Medline: 22445452]
14. Liu Q, Li M, Gao Y, Jiang T, Han B, Zhao G, et al. Effect of robotic versus open pancreaticoduodenectomy on postoperative length of hospital stay and complications for pancreatic head or periampullary tumours: a multicentre, open-label randomised controlled trial. Lancet Gastroenterol Hepatol 2024;9(5):428-437. [doi: 10.1016/S2468-1253(24)00005-0] [Medline: 38428441]
15. Dionigi R, Dionigi G, Rovera F, Boni L. Postoperative fever. Surg Infect (Larchmt) 2006;7 Suppl 2:S17-S20. [doi: 10.1089/sur.2006.7.s2-17] [Medline: 16895496]
16. Lesperance R, Lehman R, Lesperance K, Cronk D, Martin M. Early postoperative fever and the "routine" fever work-up: results of a prospective study. J Surg Res 2011;171(1):245-250. [doi: 10.1016/j.jss.2010.03.009] [Medline: 20655062]
17. Vermeulen H, Storm-Versloot MN, Goossens A, Speelman P, Legemate DA. Diagnostic accuracy of routine postoperative body temperature measurements. Clin Infect Dis 2005;40(10):1404-1410. [doi: 10.1086/429621] [Medline: 15844061]
18. Mackowiak P, Wasserman S, Levine M. A critical appraisal of 98.6 degrees F, the upper limit of the normal body temperature, and other legacies of Carl Reinhold August Wunderlich. JAMA 1992;268(12):1578-1580. [Medline: 1302471]
19. Suehiro E, Sadahiro H, Goto H, Oku T, Oka F, Fujiyama Y, et al. Importance of early postoperative body temperature management for treatment of subarachnoid hemorrhage. J Stroke Cerebrovasc Dis 2016;25(6):1482-1488. [doi: 10.1016/j.jstrokecerebrovasdis.2016.01.053] [Medline: 27020121]
20. Csendes A, Burgos AM, Roizblatt D, Garay C, Bezama P. Inflammatory response measured by body temperature, C-reactive protein and white blood cell count 1, 3, and 5 days after laparotomic or laparoscopic gastric bypass surgery. Obes Surg 2009;19(7):890-893. [doi: 10.1007/s11695-008-9702-9] [Medline: 18830782]

21. Kendrick JE, Numnum TM, Estes JM, Kimball KJ, Leath CA, Straughn JJ. Conservative management of postoperative fever in gynecologic patients undergoing major abdominal or vaginal operations. J Am Coll Surg 2008;207(3):393-397. [doi: 10.1016/j.jamcollsurg.2008.04.001] [Medline: 18722945]
22. Ward DT, Hansen EN, Takemoto SK, Bozic KJ. Cost and effectiveness of postoperative fever diagnostic evaluation in total joint arthroplasty patients. J Arthroplasty 2010;25(6 Suppl):43-48. [doi: 10.1016/j.arth.2010.03.016] [Medline: 20452174]
23. Athanassious C, Samad A, Avery A, Cohen J, Chalnick D. Evaluation of fever in the immediate postoperative period in patients who underwent total joint arthroplasty. J Arthroplasty 2011;26(8):1404-1408. [doi: 10.1016/j.arth.2011.02.019] [Medline: 21477972]
24. de la Torre SH, Mandel L, Goff BA. Evaluation of postoperative fever: usefulness and cost-effectiveness of routine workup. Am J Obstet Gynecol 2003;188(6):1642-1647. [doi: 10.1067/mob.2003.397] [Medline: 12825005]
25. Hwang CJ, Park S, Park JY, Cho JH, Lee CS, Lee DH. Sustained postoperative fever without evident cause after spine instrumentation as an indicator of surgical site infection. J Bone Joint Surg Am 2020;102(16):1434-1444. [doi: 10.2106/JBJS.19.01490] [Medline: 32590401]
26. Ji Y, Han D, Han L, Xie S, Pan S. The accuracy of a wireless axillary thermometer for core temperature monitoring in pediatric patients having noncardiac surgery: an observational study. J Perianesth Nurs 2021;36(6):685-689 [FREE Full text] [doi: 10.1016/j.jopan.2021.02.008] [Medline: 34384688]
27. Pei L, Huang Y, Mao G, Sessler DI. Axillary temperature, as recorded by the iThermonitor WT701, well represents core temperature in adults having noncardiac surgery. Anesth Analg 2018;126(3):833-838. [doi: 10.1213/ANE.0000000000002706] [Medline: 29293176]
28. Carandina S, Zulian V, Nedelcu A, Sista F, Danan M, Nedelcu M. Laparoscopic sleeve gastrectomy follow-up: use of connected devices in the postoperative period. Surg Obes Relat Dis 2019;15(7):1058-1065. [doi: 10.1016/j.soard.2019.03.033] [Medline: 31153891]
29. Harding C, Pompei F, Bordonaro SF, McGillicuddy DC, Burmistrov D, Sanchez LD. Fever incidence is much lower in the morning than the evening: Boston and US national triage data. West J Emerg Med 2020;21(4):909-917 [FREE Full text] [doi: 10.5811/westjem.2020.3.45215] [Medline: 32726264]
30. Diamond A, Lye CT, Prasad D, Abbott D. One size does not fit all: assuming the same normal body temperature for everyone is not justified. PLoS One 2021;16(2):e0245257 [FREE Full text] [doi: 10.1371/journal.pone.0245257] [Medline: 33534845]
31. Obana KK, Lin AJ, Yang J, Ryan DD, Goldstein RY, Kay RM. Fever after varus derotational osteotomy is common, but not a risk factor for infection. Medicine (Baltimore) 2020;99(1):e18613 [FREE Full text] [doi: 10.1097/MD.0000000000018613] [Medline: 31895814]
32. Petretta R, McConkey M, Slobogean GP, Handel J, Broekhuyse HM. Incidence, risk factors, and diagnostic evaluation of postoperative fever in an orthopaedic trauma population. J Orthop Trauma 2013;27(10):558-562. [doi: 10.1097/BOT.0b013e31828af4df] [Medline: 23412504]
33. Yoo JH, Kim KT, Kim TY, Hwang JH, Chang JD. Postoperative fever after hemiarthroplasty in elderly patients over 70 years of age with displaced femoral neck fracture: Necessity of routine workup? Injury 2017;48(2):441-446. [doi: 10.1016/j.injury.2016.12.013] [Medline: 28040259]
34. Kelly GS. Body temperature variability (Part 2): masking influences of body temperature variability and a review of body temperature variability in disease. Altern Med Rev 2007;12(1):49-62. [Medline: 17397267]
35. Dellinger EP. Should we measure body temperature for patients who have recently undergone surgery? Clin Infect Dis 2005 1;40(10):1411-1412. [doi: 10.1086/429629] [Medline: 15844062]
36. Storm-Versloot MN, Knops AM, Ubbink DT, Goossens A, Legemate DA, Vermeulen H. Long-term adherence to a local guideline on postoperative body temperature measurement: mixed methods analysis. J Eval Clin Pract 2012;18(4):841-847. [doi: 10.1111/j.1365-2753.2011.01687.x] [Medline: 21518400]
37. Frank SM, Kluger MJ, Kunkel SL. Elevated thermostatic setpoint in postoperative patients. Anesthesiology 2000;93(6):1426-1431 [FREE Full text] [doi: 10.1097/00000542-200012000-00014] [Medline: 11149437]
38. Karam JA, Zmistowski B, Restrepo C, Hozack WJ, Parvizi J. Fewer postoperative fevers: an unexpected benefit of multimodal pain management? Clin Orthop Relat Res 2014;472(5):1489-1495 [FREE Full text] [doi: 10.1007/s11999-014-3555-4] [Medline: 24615425]


## Abbreviations

**ERAS:** Enhanced Recovery After Surgery

*Edited by T de Azevedo Cardoso; submitted 06.07.23; peer-reviewed by S Lampridis, P Ebrahimi; comments to author 13.03.24; revised version received 05.05.24; accepted 26.06.24; published 09.10.24.*

*Please cite as:*
*Wang S, Ji G, Feng X, Huang L, Luo J, Yu P, Zheng J, Yang B, Wang X, Zhao Q*
*Temperature Measurement Timings and the Fever Detection Rate After Gastrointestinal Surgery: Retrospective Cross-Sectional Study*
*Interact J Med Res 2024;13:e50585*
*URL: https://www.i-jmr.org/2024/1/e50585*
*doi:10.2196/50585*
*PMID:*

Short Paper

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

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

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

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

## 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.

***(Interact J Med Res 2024;13:e50284)*** doi:10.2196/50284

**KEYWORDS**

French; sense of coherence; salutogenesis; SOC; Sense of Coherence scale; validation; validscale; well-being; promoting; resilience; validity; 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 of 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 α and Loevinger H coefficients were all <0.70 and <0.30, respectively (0.64 and 0.29 for comprehensibility, 0.56 and 0.26 for manageability, and 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 α 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.

| Scales and items | Missing data, % | Patients, n | Response category, % | | | | | | | Cronbach α[a] | Loevinger H |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | | |
| **SOC-13** | | | | | | | | | | | |
| **Dimension 1: comprehensibility** | | | | | | | | | | | |
| Item 2 | 1.37 | 793 | 8.45 | 19.29 | 30.77 | 22.32 | 10.09 | 6.43 | 2.65 | 0.71 | 0.10 |
| Item 6 | 1.37 | 793 | 3.28 | 7.44 | 10.34 | 16.27 | 19.17 | 27.49 | 16.02 | 0.53 | 0.35 |
| Item 8 | 1.37 | 793 | 3.28 | 7.06 | 12.11 | 19.55 | 15.51 | 25.35 | 17.15 | 0.51 | 0.37 |
| Item 9 | 2.11 | 787 | 2.03 | 5.84 | 11.56 | 18.17 | 14.49 | 25.54 | 22.36 | 0.55 | 0.33 |
| Item 11 | 2.11 | 787 | 3.30 | 8.51 | 12.33 | 22.24 | 19.44 | 23.38 | 10.80 | 0.61 | 0.26 |
| **Dimension 2: manageability** | | | | | | | | | | | |
| Item 3 | 1.74 | 790 | 8.73 | 16.58 | 29.49 | 18.10 | 12.66 | 11.27 | 3.16 | 0.57 | 0.20 |
| Item 5 | 1.37 | 793 | 3.40 | 5.80 | 11.98 | 20.18 | 17.40 | 24.46 | 16.77 | 0.50 | 0.25 |
| Item 10 | 1.87 | 789 | 3.17 | 11.15 | 15.72 | 17.49 | 13.94 | 20.66 | 17.87 | 0.43 | 0.30 |
| Item 13 | 2.11 | 787 | 2.54 | 6.35 | 10.67 | 14.23 | 16.90 | 31.77 | 17.53 | 0.46 | 0.29 |
| **Dimension 3: meaningfulness** | | | | | | | | | | | |
| Item 1 | 1.37 | 793 | 5.04 | 5.42 | 11.73 | 14.75 | 14.00 | 25.98 | 23.08 | 0.55 | 0.06 |
| Item 4 | 1.37 | 793 | 2.02 | 1.77 | 2.27 | 13.11 | 20.81 | 33.80 | 26.23 | 0.31 | 0.22 |
| Item 7 | 1.49 | 792 | 1.39 | 4.17 | 6.82 | 20.45 | 22.35 | 30.56 | 14.27 | 0.34 | 0.21 |
| Item 12 | 1.99 | 788 | 4.31 | 5.46 | 9.14 | 13.07 | 19.04 | 32.11 | 16.88 | 0.32 | 0.21 |
| **SOC-8** | | | | | | | | | | | |
| Item 6 | 1.37 | 793 | 3.28 | 7.44 | 10.34 | 16.27 | 19.17 | 27.49 | 16.02 | 0.79 | 0.40 |
| Item 8 | 1.37 | 793 | 3.28 | 7.06 | 12.11 | 19.55 | 15.51 | 25.35 | 17.15 | 0.78 | 0.43 |
| Item 9 | 2.11 | 787 | 2.03 | 5.84 | 11.56 | 18.17 | 14.49 | 25.54 | 22.36 | 0.80 | 0.37 |
| Item 11 | 2.11 | 787 | 3.30 | 8.51 | 12.33 | 22.24 | 19.44 | 23.38 | 10.80 | 0.81 | 0.34 |
| Item 10 | 1.87 | 789 | 3.17 | 11.15 | 15.72 | 17.49 | 13.94 | 20.66 | 17.87 | 0.79 | 0.41 |
| Item 13 | 2.11 | 787 | 2.54 | 6.35 | 10.67 | 14.23 | 16.90 | 31.77 | 17.53 | 0.80 | 0.38 |
| Item 4 | 1.37 | 793 | 2.02 | 1.77 | 2.27 | 13.11 | 20.81 | 33.80 | 26.23 | 0.81 | 0.33 |
| Item 12 | 1.99 | 788 | 4.31 | 5.46 | 9.14 | 13.07 | 19.04 | 32.11 | 16.88 | 0.80 | 0.35 |

[a]Cronbach α is calculated if the item is removed; for example, if item 2 of SOC-13 were removed, the Cronbach α 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.

| Scales and items | Factor loading (SE) | Intercept (SE) | Error variance | Correlation matrix | | |
|---|---|---|---|---|---|---|
| | | | | Dimension 1 | Dimension 2 | Dimension 3 |
| **SOC-13[a]** | | | | | | |
| **Dimension 1: comprehensibility (variance=0.06)** | | | | | | |
| Item 2 | 1.00 | 3.37 (0.05) | 1.95 | 0.124 | 0.320 | 0.105 |
| Item 6 | 4.29 (0.95) | 4.87 (0.06) | 1.49 | 0.520 | 0.517 | 0.377 |
| Item 8 | 4.70 (1.04) | 4.82 (0.06) | 1.32 | 0.562 | 0.552 | 0.385 |
| Item 9 | 3.71 (0.84) | 5.03 (0.06) | 1.76 | 0.476 | 0.415 | 0.298 |
| Item 11 | 3.28 (0.75) | 4.59 (0.06) | 1.80 | 0.361 | 0.422 | 0.277 |
| **Dimension 2: manageability (variance=0.17)** | | | | | | |
| Item 3 | 1.00 | 3.56 (0.06) | 2.23 | 0.340 | 0.253 | 0.164 |
| Item 5 | 1.77 (0.28) | 4.83 (0.06) | 2.10 | 0.397 | 0.339 | 0.265 |
| Item 10 | 2.73 (0.40) | 4.62 (0.06) | 1.80 | 0.566 | 0.415 | 0.340 |
| Item 13 | 2.29 (0.34) | 5.03 (0.06) | 1.65 | 0.516 | 0.390 | 0.324 |
| **Dimension 3: meaningfulness (variance=0.02)** | | | | | | |
| Item 1 | 1.00 | 4.97 (0.06) | 3.06 | 0.047 | 0.072 | 0.087 |
| Item 4 | 5.72 (3.13) | 5.55 (0.05) | 1.17 | 0.383 | 0.305 | 0.324 |
| Item 7 | 4.27 (2.34) | 5.07 (0.05) | 1.62 | 0.289 | 0.250 | 0.299 |
| Item 12 | 7.51 (4.10) | 5.00 (0.06) | 1.63 | 0.430 | 0.434 | 0.317 |
| **SOC-8[b] (variance=1.13)** | | | | | | |
| Item 6 | 1.00 | 4.87 (0.06) | 1.54 | —[c] | — | — |
| Item 8 | 1.12 (0.07) | 4.82 (0.06) | 1.32 | — | — | — |
| Item 9 | 0.91 (0.07) | 5.03 (0.06) | 1.71 | — | — | — |
| Item 11 | 0.79 (0.06) | 4.58 (0.06) | 1.80 | — | — | — |
| Item 10 | 1.09 (0.07) | 4.62 (0.06) | 1.73 | — | — | — |
| Item 13 | 0.91 (0.06) | 5.02 (0.06) | 1.61 | — | — | — |
| Item 4 | 0.63 (0.05) | 5.55 (0.05) | 1.31 | — | — | — |
| Item 12 | 0.84 (0.07) | 5.01 (0.06) | 1.87 | — | — | — |

[a]SOC-13 scale: $\chi^2_{62}$=436.36, $\chi^2/df$=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.

[b]Unidimensional SOC-8 scale: $\chi^2_{20}$=61.90, $\chi^2/df$=3.1, root mean square error of approximation=0.052, standardized root mean square residual=0.030, comparative fit index=0.973.

[c]Not applicable; the dimensions are only relevant to the SOC-13 scale.

The results were similar for the unidimensional SOC-13, except that Cronbach α was higher than that found for the 3-dimension model (Cronbach α=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 α=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 α 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-dimesion 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.
[DOCX File , 20 KB - ijmr_v13i1e50284_app1.docx ]

## References

1. Antonovsky A. Unraveling the Mystery of Health: How People Manage Stress and Stay Well. San Francisco, CA: Jossey-Bass; 1987.
2. Mittelmark M, Bauer G. The meanings of salutogenesis. In: Mittelmark MB, Sagy S, Eriksson M, Bauer GF, Pelikan JM JM, Lindström B, et al, editors. The Handbook of Salutogenesis. Cham: Springer International Publishing; 2017:7-13.
3. Eriksson M, Lindström B. Antonovsky's sense of coherence scale and the relation with health: a systematic review. J Epidemiol Community Health 2006 May;60(5):376-381 [FREE Full text] [doi: 10.1136/jech.2005.041616] [Medline: 16614325]
4. Eriksson M, Lindström B. Validity of Antonovsky's sense of coherence scale: a systematic review. J Epidemiol Community Health 2005 Jun;59(6):460-466 [FREE Full text] [doi: 10.1136/jech.2003.018085] [Medline: 15911640]
5. Gautier N. Les effets du sens de la cohérence sur la santé?: une revue systématique de la littérature. Master's thesis. DUMAS. 2020 Dec 03. URL: https://dumas.ccsd.cnrs.fr/dumas-03390364 [accessed 2024-06-11]
6. Gana K, Garnier S. Latent structure of the sense of coherence scale in a French sample. Pers Indiv Diff 2001 Nov;31(7):1079-1090. [doi: 10.1016/s0191-8869(00)00205-1]
7. Perrot B, Bataille E, Hardouin J. validscale: a command to validate measurement scales. Stata J 2018 Mar 01;18(1):29-50. [doi: 10.1177/1536867x1801800104]
8. DeVellis R. Scale Development: Theory and Applications. Fourth edition. Los Angeles, CA: SAGE; 2017.
9. Nunnally J. Psychometric Theory. New York, NY: McGraw-Hill; 1978.
10. Loevinger J. The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. Psychol Bull 1948 Nov;45(6):507-529. [doi: 10.1037/h0055827] [Medline: 18893224]

11. Gadelrab H. Evaluating the Fit of Structural Equation Models: Sensitivity to Specification Error and Descriptive Goodness-of-Fit Indices. Saarbrucken, Germany: LAP LAMBERT Academic Publishing; 2010.
12. Gunawan J, Marzilli C, Aungsuroch Y. Establishing appropriate sample size for developing and validating a questionnaire in nursing research. Belitung Nurs J 2021;7(5):356-360 [FREE Full text] [doi: 10.33546/bnj.1927] [Medline: 37496511]
13. Quintard B, Constant A, Lakdja F, Labeyrie-Lagardère H. Factors predicting sexual functioning in patients 3 months after surgical procedures for breast cancer: the role of the sense of coherence. Eur J Oncol Nurs 2014 Feb;18(1):41-45. [doi: 10.1016/j.ejon.2013.09.008] [Medline: 24169031]
14. Boeckxstaens P, Vaes B, De Sutter A, Aujoulat I, van Pottelbergh G, Matheï C, et al. A high sense of coherence as protection against adverse health outcomes in patients aged 80 years and older. Ann Fam Med 2016 Jul;14(4):337-343 [FREE Full text] [doi: 10.1370/afm.1950] [Medline: 27401421]
15. Sebo P. Validation of the sense of coherence (SOC) scale in a French population. Open Science Framework. URL: https://osf.io/pj5mr/ [accessed 2024-06-11]


## 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


*Edited by T de Azevedo Cardoso; submitted 26.06.23; peer-reviewed by H Hu, S Sun; comments to author 19.01.24; revised version received 28.01.24; accepted 11.03.24; published 16.07.24.*

*Please cite as:*
*Sebo P, Tudrej B, Bernard A, Delaunay B, Dupuy A, Malavergne C, Maisonneuve H*
*Validation and Refinement of the Sense of Coherence Scale for a French Population: Observational Study*
*Interact J Med Res 2024;13:e50284*
*URL: https://www.i-jmr.org/2024/1/e50284*
*doi:10.2196/50284*
*PMID:39012689*

Review

# Gamification and Oral Health in Children and Adolescents: Scoping Review

Rui Moreira[1,2*], MD; Augusta Silveira[3,4,5*], MSc, DMD, PhD; Teresa Sequeira[3,4,5*], PhD; Nuno Durão[2*], MD; Jessica Lourenço[2*], MD; Inês Cascais[6*], MD; Rita Maria Cabral[2*], MD; Tiago Taveira Gomes[3,7*], MSc, ME, MD, PhD

[1]School of Medicine and Biomedical Sciences, Instituto de Ciencias Biomédicas Abel Salazar, University of Porto, Porto, Portugal

[2]Department of Stomatology and Oral Maxillofacial Surgery, University Hospital Center of Santo António, Porto, Portugal

[3]Faculty of Health Sciences, Fernando Pessoa University, University of Porto, Porto, Portugal

[4]Institute for Research, Innovation and Development, Fernando Pessoa Foundation, University of Porto, Porto, Portugal

[5]Center for Innovative Biomedicine and Biotechnology, University of Coimbra, Coimbra, Portugal

[6]Paediatrics Department, Maternal & Child Center of the North, University Hospital Center of Santo António, Porto, Portugal

[7]Department 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.

***(Interact J Med Res 2024;13:e35132)*** doi:10.2196/35132

**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.
- (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, gray 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.

The search string used on PubMed was “Gamification” [mh] OR “oral healthcare” [tiab] OR “gamification” [tiab] OR ad [tiab] OR “applications” [tiab] OR “gamification” [tiab] OR “digital” [tiab] OR “games” [tiab] OR “outcomes” [mh] OR policies [tiab] OR gamification [tiab] OR “oral health” [mh:noexp] (see Multimedia Appendix 1).

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 prepiloted 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.

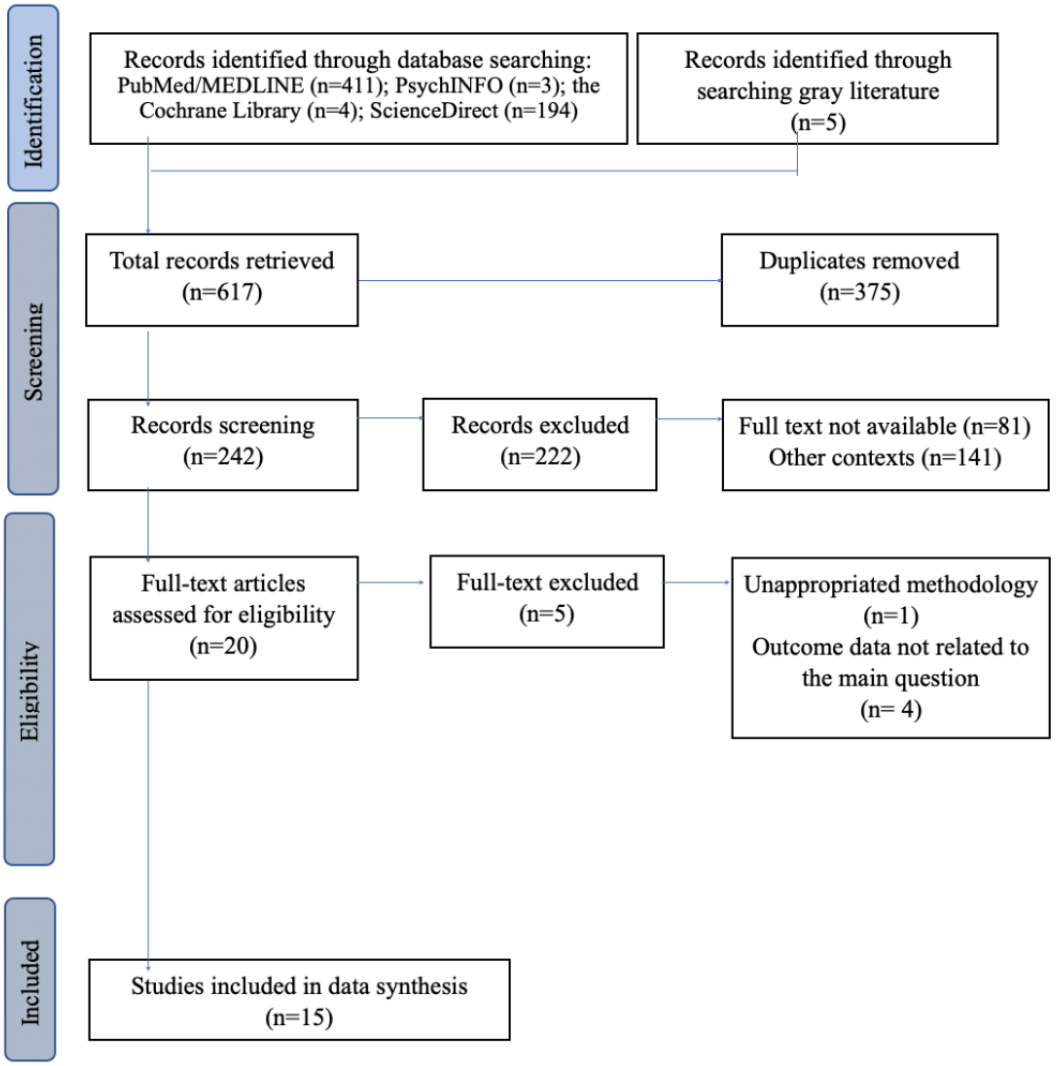
Data from the scoping literature were extracted into an Excel (Microsoft Corp) data sheet using a support checklist. The data sheet was divided into sections dedicated to a theory, area, concept, theme, or element from the framework of gamification and oral health care, including game mode, population 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.

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



## 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 [9,27,60-62,64,65]. One-time cross-sectional questionnaires were applied in 3 studies [26,66,67].

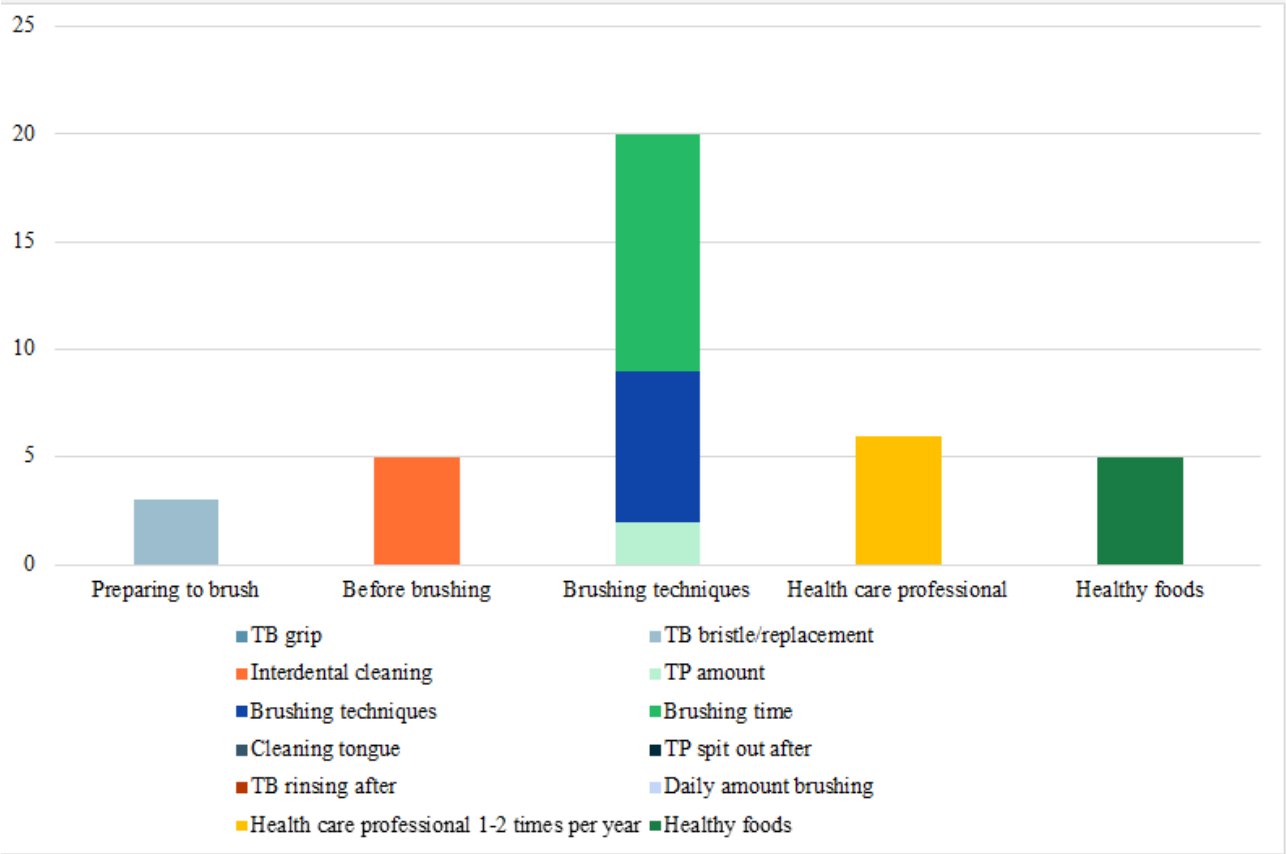
The sample size in the selected studies ranged from 34 to 190 individuals. All participants were children younger than 13 years, except for 4 studies that also included adults [26,61,62,65]. Mobile apps were the predominant modality used to change oral health care behavior in the studies (n=12), with 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).

| Behavior change techniques | Value, n (%) |
|---|---|
| 1. Provide information behavior health link | 10 (7.3) |
| 2. Provide information on consequences | 8 (6.6) |
| 3. Provide information seeking others' approval | 1 (0.7) |
| 4. Prompt intention formation | 11 (8) |
| 5. Prompt barrier identification | 3 (2.2) |
| 6. Provide general encouragement | 7 (5.1) |
| 7. Set graded tasks | 2 (1.5) |
| 8. Provide instruction | 11 (8) |
| 9. Model or demonstrate behavior | 9 (6.6) |
| 10. Prompt specific goal setting | 4 (2.9) |
| 11. Prompt review of the behavioral goals | 5 (3.6) |
| 12. Prompt self-monitoring behavior | 2 (1.5) |
| 13. Provide feedback on performance | 8 (5.8) |
| 14. Provide contingent rewards | 8 (5.8) |
| 15. Teach to use prompts or cues | 4 (2.9) |
| 16. Agree on behavioral contract | 7 (5.1) |
| 17. Prompt practice | 7 (5.1) |
| 18. Use follow-up prompts | 4 (2.9) |
| 19. Provide social comparison | 2 (1.5) |
| 20. Plan social support or change | 2 (1.5) |
| 21. Prompt identification as a role model | 2 (1.5) |
| 22. Prompt self-talk | 2 (1.5) |
| 23. Relapse prevention | 4 (2.9) |
| 24. Stress management | 1 (0.7) |
| 25. Motivational interviewing | 5 (3.6) |
| 26. Time management | 7 (5.1) |

## 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.

| Game design elements | Value, n (%) |
|---|---|
| **System design (n=36)** | |
| Visual feedback | 9 (25) |
| Audible feedback | 4 (11) |
| Reminder | 5 (14) |
| Meaning | 10 (28) |
| Integration concepts | 3 (8) |
| Visually resembling games | 1 (3) |
| Fantasy | 4 (11) |
| **Challenges (n=16)** | |
| Goals | 9 (56) |
| Time pressure | 4 (25) |
| Progressive disclosure | 3 (19) |
| **Rewards (n=22)** | |
| Ownership | 9 (41) |
| Achievement | 5 (23) |
| Point system | 4 (18) |
| Badges | 0 (0) |
| Bonus | 4 (18) |
| **Social influences (n=19)** | |
| Loss aversion | 2 (11) |
| Status | 2 (11) |
| Collaboration | 3 (16) |
| Reputation | 3 (16) |
| Competition | 1 (5) |
| Envy | 1 (5) |
| Shadowing | 2 (11) |
| Social facilitation | 2 (11) |
| Conforming behavior | 0 (0) |
| Leaderboards | 2 (11) |
| Altruism | 1 (5) |
| Virtual goods | 0 (0) |
| **User specifics (n=12)** | |
| User levels | 1 (8) |
| Ideological incentives | 10 (83) |
| Virtual characters | 1 (8) |
| Self-expression | 0 (0) |

# 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 feature 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 Alkadhimi [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.

Multimedia Appendix 1
Search string.
[PNG File , 248 KB - ijmr_v13i1e35132_app1.png ]

Multimedia Appendix 2
Full paper details.
[DOCX File , 29 KB - ijmr_v13i1e35132_app2.docx ]

Multimedia Appendix 3
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 54 KB - ijmr_v13i1e35132_app3.pdf ]

## References

1. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
2. Stevens G, Mascarenhas M, Mathers C. Global health risks: progress and challenges. Bull World Health Organ 2009 Sep;87(9):646 [FREE Full text] [doi: 10.2471/blt.09.070565] [Medline: 19784438]
3. Glanz K, Rimer BK, Viswanath K. Introduction to health behavior theories that focus on individuals. In: Health Behavior and Health Education: Theory, Research, and Practice, 4th Edition. San Francisco, CA: Jossey-Bass; Jul 2008:67-74.
4. Schroeder SA. Shattuck Lecture. We can do better—improving the health of the American people. N Engl J Med 2007 Sep 20;357(12):1221-1228. [doi: 10.1056/NEJMsa073350] [Medline: 17881753]
5. Oral health: fact sheet. World Health Organization. 2023 Mar 20. URL: https://www.who.int/news-room/fact-sheets/detail/oral-health [accessed 2020-06-24]
6. Petersen PE. The World Oral Health Report 2003: continuous improvement of oral health in the 21st century—the approach of the WHO Global Oral Health Programme. Community Dent Oral Epidemiol 2003 Dec;31(Suppl 1):3-23. [doi: 10.1046/j..2003.com122.x] [Medline: 15015736]

7. Nunn JH. The burden of oral ill health for children. Arch Dis Child 2006 Mar;91(3):251-253 [FREE Full text] [doi: 10.1136/adc.2005.077016] [Medline: 16492889]
8. Rabiei S, Mohebbi SZ, Patja K, Virtanen JI. Physicians' knowledge of and adherence to improving oral health. BMC Public Health 2012 Oct 09;12:855 [FREE Full text] [doi: 10.1186/1471-2458-12-855] [Medline: 23046660]
9. Fijačko N, Gosak L, Cilar L, Novšak A, Creber RM, Skok P, et al. The effects of gamification and oral self-care on oral hygiene in children: systematic search in app stores and evaluation of apps. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e16365 [FREE Full text] [doi: 10.2196/16365] [Medline: 32673235]
10. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
11. Political declaration of the high-level meeting of the general assembly on the prevention and control of non-communicable diseases: draft resolution / submitted by the President of the General Assembly. UN General Assembly. New York: United Nations URL: https://digitallibrary.un.org/record/710899 [accessed 2023-01-24]
12. Porter ME, Teisberg EO. How physicians can change the future of health care. JAMA 2007 Mar 14;297(10):1103-1111. [doi: 10.1001/jama.297.10.1103] [Medline: 17356031]
13. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
14. Reeve J. Understanding Motivation and Emotion, Sixth Edition. New York: Wiley; Nov 3, 2014.
15. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
16. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
17. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
18. Panic K, Cauberghe V, de Pelsmacker P. Promoting dental hygiene to children: comparing traditional and interactive media following threat appeals. J Health Commun 2014;19(5):561-576. [doi: 10.1080/10810730.2013.821551] [Medline: 24393019]
19. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertainment Computing 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]
20. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
21. Kharrazi H, Lu AS, Gharghabi F, Coleman W. A scoping review of health game research: past, present, and future. Games Health J 2012;1(2):153-164 [FREE Full text] [doi: 10.1089/g4h.2012.0011] [Medline: 24416638]
22. Deterding S. The lens of intrinsic skill atoms: a method for gameful design. Human–Computer Interaction 2015 May 15;30(3-4):294-335. [doi: 10.1080/07370024.2014.993471]
23. Sawyer B. Games for health. MIT Media Lab Talks. Cambridge, MA: MIT; 2014. URL: https://www.media.mit.edu/video/view/wellbeing2014-11-05 [accessed 2023-01-25]
24. Munson SA, Poole E, Perry DB, Peyton T. Gamification and health. In: Walz SP, Deterding S, editors. The Gameful World: Approaches, Issues, Applications. Cambridge, London: MIT Press; 2015:597-623.
25. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
26. Underwood B, Birdsall J, Kay E. The use of a mobile app to motivate evidence-based oral hygiene behaviour. Br Dent J 2015 Aug 28;219(4):E2. [doi: 10.1038/sj.bdj.2015.660] [Medline: 26315196]
27. Parker K, Bharmal RV, Sharif MO. The availability and characteristics of patient-focused oral hygiene apps. Br Dent J 2019 Apr;226(8):600-604. [doi: 10.1038/s41415-019-0197-7] [Medline: 31028329]
28. Annual number of mobile app downloads worldwide 2020. Statista. URL: https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/ [accessed 2021-03-27]
29. Bohn CE, McQuistan MR, McKernan SC, Askelson NM. Preferences related to the use of mobile apps as dental patient educational aids: a pilot study. J Prosthodont 2018;27(4):329-334. [doi: 10.1111/jopr.12667] [Medline: 28872732]
30. Zotti F, Dalessandri D, Salgarello S, Piancino M, Bonetti S, Visconti L, et al. Usefulness of an app in improving oral hygiene compliance in adolescent orthodontic patients. Angle Orthod 2016;86(1):101-107 [FREE Full text] [doi: 10.2319/010915-19.1] [Medline: 25799001]
31. Dubois D, Tamburrelli G. Understanding gamification mechanisms for software development. New York, NY, USA: Association for Computing Machinery; 2013 Presented at: Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering; August 18-26, 2013; Saint Petersburg, Russia p. 659-662. [doi: 10.1145/2491411.2494589]
32. Morschheuser BS, Rivera-Pelayo V, Mazarakis A, Zacharias V. Interaction and reflection with quantified self and gamification: an experimental study. J Lit Technol 2014;15(2):136-156 [FREE Full text]

33. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". New York, NY, US: Association for Computing Machinery; 2011 Presented at: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; September 28-30, 2011; Tampere, Finland p. 9-15 URL: https://dl.acm.org/doi/abs/10.1145/2181037.2181040 [doi: 10.1145/2181037.2181040]
34. Marston HR, Hall AK. Gamification: applications for health promotion and health information technology engagement. In: Novák D, Tulu B, Brendryen H, editors. Handbook of Research on Holistic Perspectives in Gamification for Clinical Practice. Hershey, PA: IGI Global; 2016:78-104.
35. Blohm I, Leimeister JM. Gamification. Wirtschaftsinf 2013;55(4):275-278. [doi: 10.1007/s11576-013-0368-0]
36. Huotari K, Hamari J. Defining gamification: a service marketing perspective. New York, NY, US: Association for Computing Machinery; 2012 Presented at: Proceeding of the 16th International Academic MindTrek Conference; October 3-5, 2012; Tampere, Finland p. 17-22. [doi: 10.1145/2393132.2393137]
37. Rigby S, Ryan RM. Glued to Games: How Video Games Draw Us in and Hold Us Spellbound. Santa Barbara, CA: Bloomsbury Academic; 2011.
38. Seaborn K, Fels DI. Gamification in theory and action: a survey. Int J Hum Comput 2015;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
39. Deci EL, Ryan RM. Motivation, personality, and development within embedded social contexts: an overview of self-determination theory. In: Ryan RM, editor. The Oxford Handbook of Human Motivation. New York: Oxford University Press; 2012:85-108.
40. Fortier MS, Duda JL, Guerin E, Teixeira PJ. Promoting physical activity: development and testing of self-determination theory-based interventions. Int J Behav Nutr Phys Act 2012;9:20 [FREE Full text] [doi: 10.1186/1479-5868-9-20] [Medline: 22385751]
41. Ng JYY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
42. Patrick H, Williams GC. Self-determination theory: its application to health behavior and complementarity with motivational interviewing. Int J Behav Nutr Phys Act 2012;9:18 [FREE Full text] [doi: 10.1186/1479-5868-9-18] [Medline: 22385676]
43. Teixeira PJ, Palmeira AL, Vansteenkiste M. The role of self-determination theory and motivational interviewing in behavioral nutrition, physical activity, and health: an introduction to the IJBNPA special series. Int J Behav Nutr Phys Act 2012;9:17 [FREE Full text] [doi: 10.1186/1479-5868-9-17] [Medline: 22385644]
44. McGonigal J. Reality Is Broken: Why Games Make Us Better and How They Can Change the World. New York: Penguin Press; 2011.
45. Nieborg DB. Am I mod or not?—an analysis of first person shooter modification culture. 2005 Presented at: Creative Gamers Seminar—Exploring Participatory Culture in Gaming; January 14-15, 2005; University of Tampere, Finland.
46. Sotamaa O. "Have fun working with our product!": critical perspectives on computer game mod competitions. 2005 Presented at: DiGRA Conference; 2005; Vancouver, British Columbia, Canada.
47. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
48. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013;106(3):76-78 [FREE Full text] [doi: 10.1177/0141076813480996] [Medline: 23481424]
49. Pereira P, Duarte E, Rebelo F, Noriega P. A review of gamification for health-related contexts. In: Marcus A, editor. Design, User Experience, and Usability: User Experience Design for Diverse Interaction Platforms and Environments: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II. Cham: Springer; 2014:742-753.
50. Sola D, Couturier J, Voyer B. Unlocking patient activation in chronic disease care. Br J Health Care Manag 2015;21(5):220-225. [doi: 10.12968/bjhc.2015.21.5.220]
51. Hamari J, Koivisto J, Sarsa H. Does gamification work?—a literature review of empirical studies on gamification. : IEEE; 2014 Presented at: 47th Hawaii International Conference on System Sciences; January 6-9, 2014; Waikoloa, HI, USA p. 3025-3034. [doi: 10.1109/hicss.2014.377]
52. Juul J. Gamification backlash roundup. The Ludologist. 2011. URL: https://www.jesperjuul.net/ludologist/2011/04/02/gamification-backlash-roundup/ [accessed 2024-01-25]
53. Kawachi I. It's all in the game—the uses of gamification to motivate behavior change. JAMA Intern Med 2017;177(11):1593-1594. [doi: 10.1001/jamainternmed.2017.4798] [Medline: 28973152]
54. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
55. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
56. Center for Health Studies and Research of the University of Coimbra. URL: https://www.uc.pt/en/ceisuc/ [accessed 2022-12-15]

57. Repositório de Instrumentos de Medição e Avaliação em Saúde. RIMAS. 2022. URL: http://rimas.uc.pt/ [accessed 2024-01-25]
58. Alkilzy M, Midani R, Höfer M, Splieth C. Improving toothbrushing with a smartphone app: results of a randomized controlled trial. Caries Res 2019;53(6):628-635. [doi: 10.1159/000499868] [Medline: 31132765]
59. Scheerman JFM, van Meijel B, van Empelen P, Verrips GHW, van Loveren C, Twisk JWR, et al. The effect of using a mobile application ("WhiteTeeth") on improving oral hygiene: a randomized controlled trial. Int J Dent Hyg 2020;18(1):73-83 [FREE Full text] [doi: 10.1111/idh.12415] [Medline: 31291683]
60. Jacobson D, Jacobson J, Leong T, Lourenco S, Mancl L, Chi DL. Evaluating child toothbrushing behavior changes associated with a mobile game app: a single arm pre/post pilot study. Pediatr Dent 2019;41(4):299-303 [FREE Full text] [Medline: 31439090]
61. Zolfaghari M, Shirmohammadi M, Shahhosseini H, Mokhtaran M, Mohebbi SZ. Development and evaluation of a gamified smart phone mobile health application for oral health promotion in early childhood: a randomized controlled trial. BMC Oral Health 2021;21(1):18 [FREE Full text] [doi: 10.1186/s12903-020-01374-2] [Medline: 33413304]
62. Shirmohammadi M, Razeghi S, Shamshiri AR, Mohebbi SZ. Impact of smartphone application usage by mothers in improving oral health and its determinants in early childhood: a randomised controlled trial in a paediatric dental setting. Eur Arch Paediatr Dent 2022;23(4):629-639 [FREE Full text] [doi: 10.1007/s40368-022-00731-9] [Medline: 35841512]
63. Aljafari A, Gallagher JE, Hosey MT. Can oral health education be delivered to high-caries-risk children and their parents using a computer game?—a randomised controlled trial. Int J Paediatr Dent 2017;27(6):476-485 [FREE Full text] [doi: 10.1111/ipd.12286] [Medline: 28052471]
64. Hotwani K, Sharma K, Nagpal D, Lamba G, Chaudhari P. Smartphones and tooth brushing: content analysis of the current available mobile health apps for motivation and training. Eur Arch Paediatr Dent 2020;21(1):103-108. [doi: 10.1007/s40368-019-00457-1] [Medline: 31154620]
65. Sharif MO, Alkadhimi A. Patient focused oral hygiene apps: an assessment of quality (using MARS) and knowledge content. Br Dent J 2019;227(5):383-386 [FREE Full text] [doi: 10.1038/s41415-019-0665-0] [Medline: 31520041]
66. Ito C, Marinho Filho AV, Ito M, Azevedo MM, de Almeida MA. Preliminary evaluation of a serious game for the dissemination and public awareness on preschool children's oral health. Stud Health Technol Inform 2013;192:1034. [Medline: 23920808]
67. Campos LFXA, Cavalcante JP, Machado DP, Marçal E, de B Silva PG, Rolim JPML. Development and evaluation of a mobile oral health application for preschoolers. Telemed J E Health 2019;25(6):492-498. [doi: 10.1089/tmj.2018.0034] [Medline: 30036174]
68. Aljafari A, Rice C, Gallagher JE, Hosey MT. An oral health education video game for high caries risk children: study protocol for a randomized controlled trial. Trials 2015;16:237 [FREE Full text] [doi: 10.1186/s13063-015-0754-6] [Medline: 26016586]
69. de Oliveira LBJ, Jorge MSB. Uso da gamificação com vistas à prevenção e promoção da saúde bucal: protocolo de revisão de escopo. Use of gamification for the prevention and promotion of oral health: scope review protocol. Uso de la gamificación para la prevención y promoción de la saludbucal: protocolo de revisión del alcance [Article in Portuguese]. Res Soc Dev 2022;11(9):e31811931939. [doi: 10.33448/rsd-v11i9.31939]
70. Wang K, Yu KF, Liu P, Lee GHM, Wong MCM. Can mHealth promotion for parents help to improve their children's oral health? a systematic review. J Dent 2022;123:104185 [FREE Full text] [doi: 10.1016/j.jdent.2022.104185] [Medline: 35691452]
71. Moreira R. Gamificação na promoção da saúde oral. Gamification in oral health promotion [Article in Portuguese]. In: Conclusion of MBA Advanced Course on Management of Healthcare Organizations and Services. Porto, Portugal: Repository of the University Fernando Pessoa; 2021.
72. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
73. Wang VCX. Handbook of Research on Advancing Health Education through Technology. Hershey, PA: IGI Global; 2015.
74. Zaror C, Mariño R, Atala-Acevedo C. Current state of serious games in dentistry: a scoping review. Games Health J 2021;10(2):95-108 [FREE Full text] [doi: 10.1089/g4h.2020.0042] [Medline: 33818135]
75. Figueiredo M, Garcia M, Barone DAC, Roberta OR, Lurmmerz G. Gamification in oral health: experience with school of rural area. Rev ABENO 2015;15(3):98-108 [FREE Full text]
76. Scheerman JFM, van Meijel B, van Empelen P, Kramer GJC, Verrips GHW, Pakpour AH, et al. Study protocol of a randomized controlled trial to test the effect of a smartphone application on oral-health behavior and oral hygiene in adolescents with fixed orthodontic appliances. BMC Oral Health 2018;18(1):19 [FREE Full text] [doi: 10.1186/s12903-018-0475-9] [Medline: 29415697]


## Abbreviations

**PICO:** Patient, Intervention, Comparison, Outcome

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

*Edited by T de Azevedo Cardoso; submitted 22.11.21; peer-reviewed by G Tobias, A Teles; comments to author 03.02.22; revised version received 30.04.22; accepted 08.11.23; published 04.04.24.*

*Please cite as:*
*Moreira R, Silveira A, Sequeira T, Durão N, Lourenço J, Cascais I, Cabral RM, Taveira Gomes T*
*Gamification and Oral Health in Children and Adolescents: Scoping Review*
*Interact J Med Res 2024;13:e35132*
*URL: https://www.i-jmr.org/2024/1/e35132*
*doi:10.2196/35132*
*PMID:38573750*

Original Paper

# Dropout in a Longitudinal Survey of Amazon Mechanical Turk Workers With Low Back Pain: Observational Study

Nabeel Qureshi[1], MPH, MPhil; Ron D Hays[1,2], PhD; Patricia M Herman[1], PhD, ND

[1]RAND Health Care, RAND Corporation, Santa Monica, CA, United States

[2]Division of General Internal Medicine and Health Services Research, Department of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

**Corresponding Author:**
Nabeel Qureshi, MPH, MPhil
RAND Health Care
RAND Corporation
1776 Main Street
Santa Monica, CA, 90401
United States
Phone: 1 3103930411 ext 6054
Email: nqureshi@rand.org

## Abstract

**Background:** Surveys of internet panels such as Amazon's Mechanical Turk (MTurk) are common in health research. Nonresponse in longitudinal studies can limit inferences about change over time.

**Objective:** This study aimed to (1) describe the patterns of survey responses and nonresponse among MTurk members with back pain, (2) identify factors associated with survey response over time, (3) assess the impact of nonresponse on sample characteristics, and (4) assess how well inverse probability weighting can account for differences in sample composition.

**Methods:** We surveyed adult MTurk workers who identified as having back pain. We report participation trends over 3 survey waves and use stepwise logistic regression to identify factors related to survey participation in successive waves.

**Results:** A total of 1678 adults participated in wave 1. Of those, 983 (59%) participated in wave 2 and 703 (42%) in wave 3. Participants who did not drop out took less time to complete previous surveys (30 min vs 35 min in wave 1, $P<.001$; 24 min vs 26 min in wave 2, $P=.02$) and reported having fewer health conditions (5.88 vs 6.6, $P<.001$). In multivariate models predicting responding at wave 2, lower odds of participation were associated with more time to complete the baseline survey (odds ratio [OR] 0.98, 95% CI 0.97-0.99), being Hispanic (compared with non-Hispanic, OR 0.69, 95% CI 0.49-0.96), having a bachelor's degree as their terminal degree (compared with all other levels of education, OR 0.58, 95% CI 0.46-0.73), having more pain interference and intensity (OR 0.75, 95% CI 0.64-0.89), and having more health conditions. In contrast, older respondents (older than 45 years age compared with 18-24 years age) were more likely to respond to the wave 2 survey (OR 2.63 and 3.79, respectively) and those whose marital status was divorced (OR 1.81) and separated (OR 1.77) were also more likely to respond to the wave 2 survey. Weighted analysis showed slight differences in sample demographics and conditions and larger differences in pain assessments, particularly for those who responded to wave 2.

**Conclusions:** Longitudinal studies on MTurk have large, differential dropouts between waves. This study provided information about the individuals more likely to drop out over time, which can help researchers prepare for future surveys.

***(Interact J Med Res 2024;13:e58771)*** doi:10.2196/58771

**KEYWORDS**

chronic low back pain; Mechanical Turk; MTurk; survey attrition; survey weights; Amazon; occupational health; manual labor

## Introduction

### Background

Conducting surveys on platforms such as Amazon Mechanical Turk (MTurk) have proliferated as a cost-effective and fast way of collecting data about health [1-3]. The number of studies using MTurk for social science research has been steadily increasing due in part to the ease of use, existing tools to support research activities, and quick turnaround for data collection [4]. In addition to the relatively low costs of conducting survey research with MTurk [5], another potential benefit is being able

to reach participants and retain them in longitudinal studies [6]. If the goals of a research study involve having a representative sample of participants, it is important to assess how well MTurk can meet that need.

MTurk is one of many ways to collect nonprobability survey samples that are defined and created by researchers from a pool of available participants [7,8]. Previous research has found differences between the characteristics of MTurk respondents and the US general population. MTurk participants are generally younger, more likely to be female, White, have lower income, and have higher education levels compared with the US general population, differences that have persisted over time [9-14].

## Previous Work

Collecting a nonprobability versus a probability-based sample may depend on the research question. While "statistical sampling theory suggests that any estimate of a parameter will be more accurate when that parameter is estimated using data from a random sample" [7], adjustment approaches after sample collection may improve the comparability of a nonprobability sample to the general population [15]. However, nonresponse bias due to attrition in samples can significantly impact inferences drawn from either a probability or a nonprobability panel [16,17]. Attrition over time can reduce sample size, which lowers the power of any statistical analysis, while differential attrition can bias inference in less predictable ways [7]. Several methods exist to control for bias introduced by nonresponse, including sample weighting, that reduce the impact of nonresponse on inferences. Survey attrition has been noted as a critical concern with using MTurk [4]. Still, there is limited information about the effects of survey attrition in longitudinal studies using MTurk and the extent to which it limits the inferences that can be drawn [18].

Previous research has shown that nonresponse patterns vary by survey population and survey type in MTurk. In a 3-wave longitudinal study fielded from April 2020 to March 2021, MTurk respondents who were younger, Hispanic, and had self-rated difficulty with the survey were more likely to drop out in subsequent survey waves [19]. Rates of nonresponse for short-term studies (ie, a few days to a few weeks) tend to be lower than for long-term studies (ie, a month or more) [20]. Factors related to nonresponse vary by survey type, time between survey waves, and the underlying population [21-23]. Surveys that are longer and with greater response burden produce higher rates of nonresponse [24] in all types of longitudinal surveys [25,26], including internet survey panels [22,23]. Most of these studies have focused primarily on samples of the general population [19] rather than on subgroups with clinical conditions.

We use the Mercer et al [27] framework to assess the impact of nonresponse on estimation and bias in a longitudinal study of individuals with back pain. The authors propose a 3-element assessment to assess the impact of selection bias in survey estimates. We adapt this framework to and evaluate nonresponse, assuming the baseline data reflect the population and that nonresponse bias is similar to selection bias when assessing longitudinal surveys. The 3 elements proposed by Mercer et al [27] include "exchangeability" (whether all confounding variables are known and measured), "positivity" (whether the sample includes all necessary kinds of units in the target population), and "composition" (does the sample distribution match the target population concerning confounders, or can it be adjusted to match the target population). Assessing and addressing issues with exchangeability, positivity, and composition have been shown to improve inference in causal analysis and survey analysis to deal with selection bias issues. In this article, we use the same framework to improve inference from nonresponse bias in MTurk studies.

It is essential to understand the factors associated with attrition in longitudinal studies with internet panels, given their widespread use. To improve exchangeability, it is also important to understand and assess what factors could confound inference due to nonresponse. While studies have previously examined these issues among general populations, the factors associated with attrition may vary among populations with different health conditions. It is estimated that 39% of the US adult population has back pain [28]; back pain accounts for the largest share of years lived with disability in the United States [29]. Healthier individuals are more likely to respond to surveys, and longitudinal surveys risk losing an increasing number of less healthy participants in successive survey waves [30,31]. Given that health and pain are multidimensional, multiple measures of health and pain may be necessary to capture the confounding due to poorer health and increased pain. As more studies use surveys to assess back pain, nonresponse due to poorer health can significantly impact inference drawn from analyses, even longitudinal analyses, if differential attrition by pain status is observed. In addition, MTurk workers are known to have a high turnover rate [32]. The inability to follow up could be another important source of attrition.

## Goals of the Study

As a part of a more extensive study, we collected survey data on MTurk from individuals who self-identify as having back pain. To improve sample quality, we implemented a range of tactics to screen out poor-quality data, requiring that participants had completed several previous tasks and met an approval threshold, as well as postsurvey data cleaning to screen out those who reported having one or both of 2 fake health conditions included on the survey. What was left was a sample of self-selected, higher-quality participants who were surveyed 3 times over 6 months.

Because of the prevalence of individuals with back pain, attention to them, and the use of survey methods to assess their back pain, we analyzed data from a 6-month 3-wave longitudinal panel survey to (1) describe the patterns of survey responses and nonresponse among MTurk members with back pain, (2) identify factors associated with survey response over time (to assess "exchangeability"), (3) assess the impact of nonresponse on sample characteristics (to assess "positivity"), and (4) assess how well inverse probability weighting can account for differences in sample composition (to assess "composition"). We hypothesize that those with poorer health, more pain symptoms or severity, specific pain, and nonchronic pain will be least likely to respond to follow-up surveys. Weighting may

be able to adjust to correct for nonresponse, but whether the sample is sufficiently varied is unclear.

## Methods

### Recruitment

We developed web-based surveys to collect data from MTurk participants and used the platform CloudResearch (formerly TurkPrime; Amazon) to field the survey in 2021 [33]. Individuals who reported having back pain at baseline (wave 1) were provided the opportunity to complete follow-up surveys after 3 months (wave 2) and 6 months (wave 3). We did not note in the wave 1 survey instructions that this was a longitudinal study because only those who met the inclusion criteria for the longitudinal study were asked if they wanted to participate in follow-up surveys. At the beginning of wave 2 and wave 3 recruitment, all eligible participants who consented to participate in follow-up survey waves were sent a recruitment email telling them the follow-up survey was available, that it would take approximately 25 minutes to complete, the payment for completing it, and that they had up to 5 weeks to return it. Weekly reminder emails (1-4 weeks after the recruitment email) were sent to all nonparticipants reiterating the timeline for survey completion, the approximate time to complete it, and the payment for completion.

Based on previous data collection efforts, we recruited individuals to have a final wave 1 sample of about 1500 individuals with back pain [9]. Those invited to participate at baseline had to have completed a minimum of 500 previous human intelligence tasks (HITs) on MTurk with a successful completion rate of at least 95%. No additional requirements were given to participate in the wave 1 survey. These threshold values were selected to enhance data quality. Previous research [34] and pilot tests of the survey found a 95% approval threshold and at least 500 completed HITs improve data quality and that limiting samples to ensure data quality does not limit the pool of available workers enough to restrict the sample to below the 1500-participant target [35]. While more recent studies [36] have shown that the approval rate is insufficient to ensure high-quality responses, we used a range of steps to ensure high-quality responses, including reputation, number of previous tasks, and attention checks (described in the Measures section). Given the structure of the MTurk interface, we are unable to determine the impact of the approval and completed HIT thresholds have on the sample profile (ie, we cannot quantify the number of individuals who tried to complete the survey but could not because of the thresholds for participation, as those individuals would not see the survey). Additional detail on data collection of wave 1 data is described by Qureshi et al [13].

### Ethical Considerations

All participants provided electronic consent at the beginning of the survey. Those who completed general health and back pain surveys at wave 1 were offered US $3.50 for their participation. Participants were offered an additional US $5 per subsequent completed survey (wave 2 and wave 3). All baseline participants (even those who did not participate in wave 2) were asked to participate in wave 3. Data were deidentified and are stored online [37]. All procedures were reviewed and approved by the research team's institutional review board (RAND Human Subjects Research Committee FWA00003425; IRB00000051) and conforming to the Declaration of Helsinki principles. The study was funded by the National Institutes of Health or the National Center for Complementary and Integrative Health (Grant 1R01AT010402).

### Measures

The main outcome variable was participation in wave 2 and waves 2 and 3, defined as a binary outcome (0 if no participation and 1 if participation). We used several exposure variables, including self-reported demographic variables, self-reported health conditions, and self-reported back pain assessments.

Each survey asked about demographic characteristics (age, sex, race or ethnicity, employment status, income, education, and marital status) and health conditions. Health conditions were assessed in 2 forms. First, we asked "Have you EVER been told by a doctor or other health professional that you had…" for each of the following conditions: hypertension, high cholesterol, heart disease, angina, heart attack, stroke, asthma, cancer, diabetes, chronic obstructive pulmonary disease, arthritis, anxiety disorder, and depression. Then, we asked "Do you currently have…" for each of the following conditions: allergies or sinus trouble, back pain, sciatica, neck pain, trouble seeing, dermatitis, stomach trouble, trouble hearing, and trouble sleeping. We included these various measures of health to allow for the examination across various dimensions of health to support "exchangeability" for inference.

We also included 2 fake conditions in the survey that were used to screen out low-quality respondents. Individuals who endorsed one or both fake conditions were not asked to participate in the back pain follow-up survey if they endorsed having back pain. Overall, 15% (996/6832) of respondents endorsed one of these fake conditions, and their responses were believed to be dishonest or careless. Those reporting fake conditions were more likely to identify as male, non-White, to be younger, report more health conditions, and take longer to complete the survey. Their responses had less internal consistency reliability on several health measures than those who did not endorse a fake condition (Hays et al [38]).

Those who reported having back pain were asked to participate in a follow-up survey that included additional questions related to their back pain. If an individual opted not to continue, they would be paid for completing the first part of the survey and were not included in further analysis. The survey included questions about whether the respondent's back pain was "chronic" according to 1 of 4 definitions (either that their back pain persisted for least 3 months, that their back pain persisted for at least 3 months, and they had pain at least half the days in the past 6 months, that a health provider told them that their pain is chronic, or that they believe their back pain is chronic). We also asked whether their back pain was due to a "specific" medical condition. We categorized individuals with back pain into 4 groups [39]—those with specific chronic back pain, those with specific nonchronic back pain, those with nonspecific chronic back pain, and those with nonspecific nonchronic pain. The survey also included the Impact Stratification Score (ISS) [40], Oswestry Disability Index (ODI) [41], Roland Morris

Disability Questionnaire (RMDQ) [42], the Pain, Enjoyment of Life and General Activity scale (PEG) [43], and the Keele STarT Back Screening Tool (SBST) [44].

### Statistical Analysis

We report response rates to the wave 2 and 3 surveys among those responding to the wave 1 survey to assess the "positivity" of samples for inference. In addition, we report descriptive statistics on age, sex, race or ethnicity, income, education, marital status, self-reported health conditions, the proportion who endorsed back pain types, and back pain measure scores for those who participated in each survey wave. We report differences for those who did and did not complete the wave 2 survey and both the wave 2 and wave 3 surveys using *t* tests for continuous and chi-square tests for categorical variables.

Next, we report estimates from stepwise logistic regression models predicting response to wave 2 (model 1) and both waves 2 and 3 (model 2). We used a backward elimination with a selection criterion of $\alpha$=.157 and a forward selection criterion of $\alpha$=.05 to select the variables to include in the models [45]. These selection criteria determine whether a variable is included in the final model. Using a backward elimination with a selection criterion of $\alpha$=.157 rather than $\alpha$=.05 is meant to reduce overfitting of the final model, a common issue associated with stepwise models [46]. We report the odds of completing the subsequent surveys. Based on previous studies, age, sex, race, and ethnicity were included in the regression models [19,47]. We also examined education, marital status, income categories, employment, health conditions, type of pain, pain impact, and time to complete the questionnaire as predictor variables.

Finally, we used inverse probability score weighting to examine sample characteristics in waves 2 and 3 based on model 1 and model 2 results to assess how well the sample weights correct for nonresponse from in later waves to assess "composition." Model weights are derived from estimated probabilities of completion using the aforementioned stepwise logistic regression models. By using inverse probability weights, we overweight respondents like those who drop out, approximating how the original sample would have looked if everyone responded to both follow-up waves. We included all candidate variables without backward elimination as a sensitivity analysis to derive inverse probability weights. Similar baseline characteristics between the full sample at baseline and weighted estimates for those who participated in later waves is an indication that observed variables can account for the level of bias introduced by sample attrition. All analyses were conducted using Stata software version MP17 (StataCorp) [48]. The study confirms to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for cohort studies (Table S1 in Multimedia Appendix 1).

## Results

### Sample Characteristics

A total of 1678 adults who responded in wave 1 qualified to take subsequent surveys, that is, did not endorse a fake condition on the wave 1 survey and consented to participate in a future survey [13]. Of those who qualified to participate from the total sample in wave 1, 983 (59%) responded in wave 2. Of the 983 who responded in wave 2, a total of 703 (42% of wave 1 respondents) also responded in wave 3. The 8 respondents who only responded in waves 1 and 3 (ie, not in wave 2) were excluded from further analyses. Compared with those who did not respond, respondents in wave 2 were older, with higher income, more likely to never have been married, less likely to be Hispanic, less educated, and less likely to be employed full-time. We saw similar trends for those who responded in both waves 2 and 3 versus those who did not (Table 1).

Table 2 shows the overall sample distribution at wave 1 and response rates in waves 2 and 3. Generally, those who were older, those who were female, non-Hispanics, not married or living with a partner, and those at low (ie, US $0-US $39,999) and high income (more than US $60,000) were more likely to respond during waves 2 and 3 than their counterparts. These differences were more apparent when comparing wave 1 with wave 3. However, when comparing wave 3 response among those who responded in wave 2, response rates were generally 65%-75% and not systematically different by characteristic. In addition, the sample prevalence of health conditions was similar between the unweighted samples of those who participated in wave 2 only and those who participated in waves 2 and 3 (Table 2).

Those who responded in wave 2 completed the wave 1 survey in less time than those who did not respond in wave 2 (30 min vs 35 min, $P$<.001). Those who responded to the wave 3 survey also reported less time completing the wave 2 survey than those who did not respond (24 min vs 26 min, $P$=.02), similar to the time advertised to complete the survey. We found no differences in the time of day (morning, afternoon, evening, or nighttime) when the baseline survey was completed between the responders and nonresponders to the wave 2 survey and the waves 2 and 3 surveys.

Respondents in wave 2 had fewer reported health conditions than those who did not respond (5.8 vs 6.6, $P$<.001). A similar trend was observed for those who responded versus those who did not respond to both waves 2 and 3, though the effect was not significant (5.7 vs 6.0, $P$=.08). There were differences between responders and nonresponders in wave 2 for 15 conditions, with nearly all being less common for responders than nonresponders, except for arthritis, anxiety, and allergies. There were also differences between responders and nonresponders in waves 2 and 3, but for fewer (11) conditions (Table 3).

Respondents to wave 2 were less likely to have nonspecific low back pain and more likely to have chronic low back pain than those who did not respond, with similar patterns for those who did and did not respond to both the waves 2 and 3 surveys. Participants in wave 2 and in both waves 2 and 3 reported less pain intensity and pain interference, and better health on the ISS, ODI, RMDQ, PEG, and SBST measures (Table 4).

**Table 1.** Characteristics of those participating in wave 1 only versus those who also responded in wave 2 (at 3 months) and in waves 2 and 3 (at both 3 and 6 months).

| Characteristic | Responded in wave 1 only (N=695), n (%) | Responded in wave 1 and 2 only (N=983), n (%) | *P* value (wave 1 vs wave 1 and 2) | Responded in all 3 waves (N=703), n (%) | *P* value (wave 1 vs all 3 waves) |
|---|---|---|---|---|---|
| Age (years), mean (SD) | 39.13 (10.84) | 42.47 (12.01) | <.001 | 43.58 (12.14) | <.001 |
| **Age category (years)** | | | <.001 | | <.001 |
| 18-24 | 26 (3.7) | 38 (3.9) | | 24 (3.4) | |
| 25-34 | 246 (35.4) | 256 (26) | | 156 (22.2) | |
| 35-44 | 238 (34.2) | 306 (31.1) | | 225 (32) | |
| 45-54 | 107 (15.4) | 202 (20.5) | | 154 (21.9) | |
| 55-65 | 63 (9.1) | 136 (13.8) | | 107 (15.2) | |
| Older than 65 | 15 (2.2) | 45 (4.6) | | 37 (5.3) | |
| **Sex** | | | .10 | | .64 |
| Female | 337 (48.5) | 531 (54) | | 387 (55) | |
| Male | 353 (50.8) | 445 (45.3) | | 312 (44.4) | |
| Transgender | 3 (0.4) | 2 (0.2) | | 1 (0.1) | |
| Do not identify as female, male, or transgender | 2 (0.3) | 5 (0.5) | | 3 (0.4) | |
| **Race** | | | .17 | | .17 |
| White | 588 (84.6) | 810 (82.4) | | 576 (81.9) | |
| Black | 59 (8.5) | 73 (7.43) | | 47 (6.7) | |
| Asian | 28 (4) | 56 (5.7) | | 47 (6.7) | |
| Native Hawaiian or Pacific Islander | 1 (0.1) | 0 (0) | | 0 (0) | |
| American Indian or Native Alaskan | 0 (0) | 0 (0) | | 0 (0) | |
| Other | 4 (0.6) | 5 (0.51) | | 3 (0.4) | |
| Multiracial | 14 (2) | 34 (3.46) | | 27 (3.8) | |
| **Ethnicity** | | | <.001 | | <.001 |
| Not Hispanic or Latin | 552 (79.4) | 901 (91.7) | | 662 (94.2) | |
| Hispanic or Latino | 143 (20.6) | 82 (8.3) | | 41 (5.8) | |
| **Education** | | | <.001 | | <.001 |
| No high school diploma | 1 (0.1) | 3 (0.3) | | 3 (0.4) | |
| High school graduate | 33 (4.8) | 98 (10) | | 78 (11.1) | |
| Some college, no degree | 85 (12.3) | 209 (21.3) | | 152 (21.6) | |
| Occupational or technical degree | 15 (2.2) | 33 (3.4) | | 25 (3.6) | |
| Associate degree | 41 (5.9) | 106 (10.8) | | 91 (12.9) | |
| Bachelor’s degree | 402 (58) | 365 (37.2) | | 228 (32.4) | |
| Master’s degree | 106 (15.3) | 138 (14.1) | | 106 (15.1) | |
| Professional school | 8 (1.2) | 16 (1.6) | | 11 (1.6) | |
| Doctoral degree | 2 (0.3) | 12 (1.2) | | 9 (1.3) | |
| **Marital status** | | | <.001 | | .003 |
| Married or living with partner | 541 (77.8) | 579 (58.9) | | 391 (55.6) | |
| Separated | 11 (1.6) | 10 (1) | | 7 (1) | |

| Characteristic | Responded in wave 1 only (N=695), n (%) | Responded in wave 1 and 2 only (N=983), n (%) | *P* value (wave 1 vs wave 1 and 2) | Responded in all 3 waves (N=703), n (%) | *P* value (wave 1 vs all 3 waves) |
|---|---|---|---|---|---|
| Divorced | 28 (4) | 105 (10.7) | | 89 (12.7) | |
| Widowed | 3 (0.4) | 11 (1.1) | | 7 (1) | |
| Never married | 112 (16.1) | 278 (28.3) | | 209 (29.7) | |
| **Income (US $)** | | | <.001 | | .87 |
| Less than 10,000 | 22 (3.2) | 48 (4.9) | | 35 (5) | |
| 10,000-19,999 | 58 (8.3) | 80 (8.1) | | 59 (8.4) | |
| 20,000-29,999 | 89 (12.8) | 113 (11.5) | | 82 (11.7) | |
| 30,000-39,999 | 67 (9.6) | 130 (13.2) | | 99 (14.1) | |
| 40,000-49,999 | 111 (16) | 112 (11.4) | | 84 (11.9) | |
| 50,000-59,999 | 121 (17.4) | 111 (11.3) | | 75 (10.7) | |
| 60,000-79,999 | 89 (12.8) | 142 (14.4) | | 99 (14.1) | |
| 80,000-99,999 | 72 (10.4) | 107 (10.9) | | 74 (10.5) | |
| 100,000-199,999 | 63 (9.1) | 121 (12.3) | | 82 (11.7) | |
| 200,000 or more | 3 (0.4) | 19 (1.9) | | 14 (2) | |
| **Employment** | | | <.001 | | .02 |
| Full-time | 514 (74.1) | 570 (58) | | 396 (56.3) | |
| Part-time | 56 (8.1) | 123 (12.5) | | 97 (13.8) | |
| Looking for work | 26 (3.8) | 55 (5.6) | | 37 (5.3) | |
| Maternity leave | 5 (0.7) | 1 (0.1) | | 1 (0.1) | |
| Not working due to health | 21 (3) | 48 (4.9) | | 35 (5) | |
| Student | 16 (2.3) | 31 (3.2) | | 15 (2.1) | |
| Retired | 12 (1.7) | 49 (5) | | 40 (5.7) | |
| Keeping house | 27 (3.9) | 59 (6) | | 45 (6.4) | |
| Other | 17 (2.5) | 47 (4.8) | | 37 (5.3) | |

**Table 2.** Overall sample distribution and response rates by characteristic for those who responded in Wave 2 (at 3 months) and in Wave 3 (at 6 months).

| Characteristics | Wave 1 (N=1678), n (%) | Wave 2 response rate of wave 1 participants (N=983), % | Wave 3 response rate of wave 1 participants (N=703), % | Wave 3 response rate of wave 2 participants (N=703), % |
|---|---|---|---|---|
| **Age category (years)** | | | | |
| 18-24 | 64 (3.8) | 59.4 | 37.5 | 63.2 |
| 25-34 | 502 (29.9) | 51 | 31.1 | 60.9 |
| 35-44 | 544 (32.4) | 56.3 | 41.4 | 73.5 |
| 45-54 | 309 (18.4) | 65.4 | 49.8 | 76.2 |
| 55-65 | 199 (11.9) | 68.3 | 53.8 | 78.7 |
| Older than 65 | 60 (3.6) | 75 | 61.7 | 82.2 |
| **Sex** | | | | |
| Female | 868 (51.7) | 61.2 | 44.6 | 72.9 |
| Male | 798 (47.6) | 55.8 | 39.1 | 70.1 |
| Transgender | 5 (0.3) | 40 | 20 | 50 |
| Do not identify as female, male, or transgender | 7 (0.4) | 71.4 | 42.9 | 60 |
| **Race** | | | | |
| White | 1398 (83.3) | 57.9 | 41.2 | 71.1 |
| Black | 132 (7.9) | 55.3 | 35.6 | 64.4 |
| Asian | 84 (5) | 66.7 | 56 | 83.9 |
| Native Hawaiian or Pacific Islander | 1 (0.1) | 0 | 0 | 0 |
| American Indian or Native Alaskan | 0 (0) | 0 | 0 | 0 |
| Other | 9 (0.5) | 55.6 | 33.3 | 60 |
| Multiracial | 48 (2.9) | 70.8 | 56.3 | 79.4 |
| **Ethnicity** | | | | |
| Not Hispanic or Latin | 1453 (86.6) | 62 | 45.6 | 73.5 |
| Hispanic or Latino | 225 (13.4) | 36.4 | 18.2 | 50 |
| **Education** | | | | |
| No high school diploma | 4 (0.2) | 75 | 75 | 100 |
| High school graduate | 131 (7.8) | 74.8 | 59.5 | 79.6 |
| Some college, no degree | 294 (17.5) | 71.1 | 51.7 | 72.7 |
| Occupational or technical degree | 48 (2.9) | 68.8 | 52.1 | 75.8 |
| Associate degree | 147 (8.8) | 72.1 | 61.9 | 85.8 |
| Bachelor’s degree | 767 (45.7) | 47.6 | 29.7 | 62.5 |
| Master’s degree | 244 (14.5) | 56.6 | 43.4 | 76.8 |
| Professional school | 24 (1.4) | 66.7 | 45.8 | 68.8 |
| Doctoral degree | 14 (0.8) | 85.7 | 64.3 | 75 |
| **Marital status** | | | | |
| Married or living with partner | 1120 (66.7) | 51.7 | 34.9 | 67.5 |
| Separated | 21 (1.3) | 47.6 | 33.3 | 70 |
| Divorced | 133 (7.9) | 78.9 | 66.9 | 84.8 |
| Widowed | 14 (0.8) | 78.6 | 50 | 63.6 |
| Never married | 390 (23.2) | 71.3 | 53.6 | 75.2 |
| **Income (US $)** | | | | |

| Characteristics | Wave 1 (N=1678), n (%) | Wave 2 response rate of wave 1 participants (N=983), % | Wave 3 response rate of wave 1 participants (N=703), % | Wave 3 response rate of wave 2 participants (N=703), % |
|---|---|---|---|---|
| Less than 10,000 | 70 (4.2) | 68.6 | 50 | 72.9 |
| 10,000-19,999 | 138 (8.2) | 58 | 42.8 | 73.8 |
| 20,000-29,999 | 202 (12) | 55.9 | 40.6 | 72.6 |
| 30,000-39,999 | 197 (11.7) | 66 | 50.3 | 76.2 |
| 40,000-49,999 | 223 (13.3) | 50.2 | 37.7 | 75 |
| 50,000-59,999 | 232 (13.8) | 47.8 | 32.3 | 67.6 |
| 60,000-79,999 | 231 (13.8) | 61.5 | 42.9 | 69.7 |
| 80,000-99,999 | 179 (10.7) | 59.8 | 41.3 | 69.2 |
| 100,000-199,999 | 184 (11) | 65.8 | 44.6 | 67.8 |
| 200,000 or more | 22 (1.3) | 86.4 | 63.6 | 73.7 |
| **Employment** | | | | |
| Full-time | 1084 (64.6) | 52.6 | 36.5 | 69.5 |
| Part-time | 179 (10.7) | 68.7 | 54.2 | 78.9 |
| Looking for work | 81 (4.8) | 67.9 | 45.7 | 67.3 |
| Maternity leave | 6 (0.4) | 16.7 | 16.7 | 100 |
| Not working due to health | 69 (4.1) | 69.6 | 50.7 | 72.9 |
| Student | 47 (2.8) | 66 | 31.9 | 48.4 |
| Retired | 61 (3.6) | 80.3 | 65.6 | 81.6 |
| Keeping house | 86 (5.1) | 68.6 | 52.3 | 76.3 |
| Other | 64 (3.8) | 73.4 | 57.8 | 78.7 |

**Table 3.** Health conditions reported by those in the baseline (wave 1) sample who also responded in wave 2 (at 3 months) and in waves 2 and 3 (at both 3 and 6 months).

| Condition | Responded in wave 1 only (N=695), mean (SD) | Responded in wave 1 and 2 only (N=983), mean (SD) | *P* value (wave 1 vs wave 1 and 2) | Responded in all 3 waves (N=703), mean (SD) | *P* value (wave 1 vs all 3 waves) |
|---|---|---|---|---|---|
| Hypertension | 0.42 (0.49) | 0.32 (0.47) | <.001 | 0.30 (0.46) | .005 |
| High cholesterol | 0.30 (0.46) | 0.26 (0.44) | .08 | 0.27 (0.44) | .67 |
| Heart disease | 0.11 (0.32) | 0.04 (0.19) | <.001 | 0.03 (0.17) | .03 |
| Angina | 0.11 (0.31) | 0.03 (0.17) | <.001 | 0.02 (0.13) | .003 |
| Heart attack | 0.09 (0.29) | 0.03 (0.17) | <.001 | 0.02 (0.14) | .006 |
| Stroke | 0.09 (0.28) | 0.03 (0.17) | <.001 | 0.03 (0.16) | .12 |
| Asthma | 0.24 (0.42) | 0.20 (0.40) | .06 | 0.20 (0.40) | .50 |
| Cancer | 0.07 (0.26) | 0.07 (0.25) | .32 | 0.08 (0.27) | .007 |
| Diabetes | 0.24 (0.43) | 0.12 (0.32) | <.001 | 0.10 (0.31) | .03 |
| COPD[a] | 0.11 (0.31) | 0.05 (0.22) | <.001 | 0.05 (0.21) | .34 |
| Arthritis | 0.20 (0.40) | 0.27 (0.44) | .004 | 0.28 (0.45) | .045 |
| Anxiety | 0.37 (0.48) | 0.42 (0.49) | .03 | 0.41 (0.49) | .17 |
| Depression | 0.52 (0.50) | 0.47 (0.50) | .04 | 0.45 (0.50) | .02 |
| Allergies | 0.44 (0.50) | 0.52 (0.50) | .001 | 0.55 (0.50) | .005 |
| Sciatica | 0.31 (0.46) | 0.25 (0.44) | .02 | 0.26 (0.44) | .09 |
| Neck pain | 0.52 (0.50) | 0.40 (0.49) | <.001 | 0.38 (0.49) | .01 |
| Trouble seeing | 0.26 (0.44) | 0.20 (0.40) | .007 | 0.20 (0.40) | .33 |
| Dermatitis | 0.18 (0.38) | 0.16 (0.37) | .27 | 0.15 (0.36) | .25 |
| Stomach trouble | 0.36 (0.48) | 0.32 (0.47) | .07 | 0.30 (0.46) | .02 |
| Trouble hearing | 0.15 (0.36) | 0.09 (0.28) | <.001 | 0.08 (0.28) | .33 |
| Trouble sleeping | 0.50 (0.50) | 0.54 (0.50) | .11 | 0.55 (0.50) | .22 |
| Number of conditions | 6.6 (3.67) | 5.8 (2.96) | <.001 | 5.7 (2.88) | .08 |

[a]COPD: chronic obstructive pulmonary disease.

**Table 4.** Pain impact reported by those in the baseline (wave 1) sample who did not and did respond at wave 2 (at 3 months) and at waves 2 and 3 (at both 3 and 6 months).

| Pain assessment | Responded in wave 1 only (N=695) | Responded in wave 1 and 2 only (N=983) | *P* value (wave 1 vs wave 1 and 2) | Responded in all 3 waves (N=703) | *P* value (wave 1 vs all 3 waves) |
|---|---|---|---|---|---|
| Nonspecific, proportion (SD) | 0.80 (0.69) | 0.55 (0.79) | <.001 | 0.64 (0.48) | <.001 |
| Chronic, proportion (SD) | 0.84 (0.37) | 0.92 (0.26) | <.001 | 0.94 (0.24) | .016 |
| Pain intensity, *z* score (SD) | 0.85 (0.88) | 0.62 (0.89) | <.001 | 0.50 (0.78) | .002 |
| Pain interference, *z* score (SD) | 0.80 (0.69) | 0.55 (0.79) | <.001 | 0.57 (0.88) | .003 |
| Impact Stratification Score (ISS), mean (SD) | 22.07 (7.41) | 19.34 (8.57) | <.001 | 18.99 (8.6) | .02 |
| Oswestry Disability Index (ODI), mean (SD) | 26.98 (15.99) | 22.39 (15.99) | <.001 | 22.06 (16.09) | .15 |
| Roland Morris Disability Questionnaire (RMDQ), mean (SD) | 10.35 (6.63) | 8.09 (6.46) | <.001 | 7.86 (6.42) | .04 |
| Pain, Enjoyment of Life and General Activity scale (PEG), mean (SD) | 4.33 (2.08) | 3.74 (2.18) | <.001 | 3.58 (2.16) | <.001 |
| Keele STarT Back Screening Tool (SBST), mean (SD) | 4.10 (2.53) | 3.48 (2.54) | <.001 | 3.38 (2.52) | .03 |

### Response Patterns in Wave 2

Table 5 presents the odds ratios (ORs) based on the logistic regression results after stepwise selection. The models identified the factors most related to responding in the wave 2 survey (Table 5; Model 1).

Respondents with longer response times (minutes) in wave 1 were less likely to respond to the wave 2 survey than those with longer response times in wave 1 (OR 0.98, 95% CI 0.97-0.99). Respondents who were Hispanic or Latino were also less likely to participate (compared with those who were not Hispanic or Latino, OR 0.69, 95% CI 0.49-0.96). In addition, those with a bachelor's degree as their terminal degree (compared with all other education groups, OR 0.58) were less likely to respond in the following survey. Those with angina (OR 0.56), diabetes (OR 0.68), and sciatica (OR 0.72) compared with those without those conditions were less likely to respond to the wave 2 survey. Finally, those with greater pain intensity and interference (mean of the *z* scores for these measures; OR 0.75), and those with any specific back pain (OR 0.36-0.71) compared with those with nonspecific pain were less likely to respond to the following survey.

In contrast, older respondents (older than 45 years compared with 18-24 years old) were more likely to respond to the wave 2 survey (OR 2.63-3.79) and those whose marital status was divorced (OR 1.81) and separated (OR 1.77) were also more likely to respond the wave 2 survey. Those with allergies (OR 1.30) and trouble sleeping (OR 1.33) compared with those without those conditions were more likely to respond to the wave 2 survey.

**Table 5.** Results of models predicting response for wave 2 (3 months) and for waves 2 and 3 (3 and 6 months).

| Variables | Model 1—Response for wave 2, odds ratio (95% CI) | Model 2—Response for waves 2 and 3, odds ratio (95% CI) |
|---|---|---|
| Time to complete (in minutes) | 0.978[a] (0.968-0.987) | 0.978[a] 0.968-0.987) |
| **Age (years)** | | |
| 18-24 (reference) | reference | reference |
| 25-34 | 1.340 (0.729-2.464) | 0.931 (0.486-1.782) |
| 35-44 | 1.731 (0.932-3.215) | 1.573 (0.811-3.051) |
| 45-54 | 2.633[b] (1.371-5.056) | 2.192[c] (1.098-4.375) |
| 55-65 | 2.868[b] (1.434-5.736) | 2.297[c] (1.103-4.780) |
| Older than 65 | 3.785[b] (1.559-9.189) | 2.662[c] (1.094-6.478) |
| **Sex** | | |
| Male (reference) | reference | reference |
| Female | 0.959 (0.763-1.206) | 0.992 (0.786-1.252) |
| Transgender | 0.264 (0.039-1.788) | 0.199 (0.019-2.045) |
| Do not identify as female, male, or transgender | 0.985 (0.166-5.842) | 1.646 (0.289-9.376) |
| **Race** | | |
| White (reference) | reference | reference |
| Black or African American | 1.253 (0.826-1.901) | 1.104 (0.715-1.706) |
| Asian | 1.427 (0.846-2.405) | 2.044[b] (1.225-3.409) |
| Native American or Alaskan Native | 4.305 (0.438-42.31) | 1.625 (0.260-10.16) |
| Other | 0.716 (0.179-2.864) | 0.523 (0.115-2.372) |
| Multiracial | 1.376 (0.691-2.739) | 1.636 (0.851-3.143) |
| **Ethnicity** | | |
| Hispanic or Latino | 0.683[c] (0.486-0.961) | 0.540[b] (0.362-0.806) |
| **Education** | | |
| Terminal degree not bachelor's degree (reference) | reference | reference |
| Bachelor's degree | 0.575[a] (0.459-0.724) | 0.446[a] (0.346-0.575) |
| **Marital status** | | |
| Neither divorced nor never married (reference) | reference | reference |
| Divorced | 1.812[c] (1.115-2.945) | 1.894b (1.223-2.935) |
| Never married | 1.769[a] (1.318-2.375) | 1.704[a] (1.282-2.265) |
| **Employment status** | | |
| Nonstudent (reference) | ns[d] (reference) | reference |
| Student | ns (reference) | 0.403[c] (0.196-0.828) |
| **Mean of pain intensity and pain interference *z* scores** | 0.754[b] (0.639-0.889) | 0.707[a] (0.595-0.841) |
| **Type of pain** | | |
| Nonspecific pain (reference) | reference | reference |
| Specific and chronic pain | 0.712[c] (0.537-0.945) | 0.624[b] (0.477-0.816) |
| Specific and nonchronic pain | 0.355[a] (0.232-0.543) | 0.330[a] (0.209-0.520) |
| **Conditions** | | |
| Hypertension | ns (reference) | 0.773[c] (0.597-1.000) |

| Variables | Model 1—Response for wave 2, odds ratio (95% CI) | Model 2—Response for waves 2 and 3, odds ratio (95% CI) |
|---|---|---|
| Angina | 0.559[c] (0.333-0.939) | 0.370[b] (0.186-0.738) |
| Diabetes | 0.678[c] (0.495-0.929) | ns (reference) |
| Allergies | 1.301[c] (1.030-1.643) | 1.321[c] (1.048-1.665) |
| Sciatica | 0.716[c] (0.549-0.935) | ns (reference) |
| Trouble sleeping | 1.330[c] (1.052-1.682) | 1.418[b] (1.112-1.807) |

[a]*P*<.001.
[b]*P*<.01.
[c]*P*<.05.
[d]ns: nonsignificant in model.

### Response Patterns in Waves 2 and 3

We found that factors related to response in wave 2 were similar to those that related to response in waves 2 and 3 (Table 5; Model 2): time to complete, age, race, ethnicity, education, marital status, pain type, pain impact, and certain chronic conditions were associated with participation in both waves 2 and 3 of the survey.

In contrast to the model predicting response in the wave 2 survey (Model 1), those who identified as Asian (OR 2.04, 95% CI 1.23-3.41) were more likely to complete the waves 2 and 3 surveys than those who identified as White. Income was no longer significant, but being a student was associated with lower odds of responding (OR 0.40, 95% CI 0.20-0.83). Similarly, having diabetes or sciatica was no longer associated with response, but hypertension was associated with lower odds of responding (OR 0.77) compared with those without the condition.

### Description of Sample With Weighting by Nonresponse

Weights were created using the final stepwise logistic regression models shown in Table 5. To derive weights for responses in wave 2, the final model included age, sex, race, ethnicity, marital status (being divorced or never married), mean pain intensity and pain interference score, type of pain (specific and chronic pain or specific and nonchronic pain), and health conditions (angina, diabetes, allergies, sciatica, and trouble sleeping). To derive weights for response in waves 2 and 3, the final model included age, sex, race, ethnicity, marital status (being divorced or never married), employment (student or not student), mean pain intensity and pain interference score, type of pain (specific and chronic pain or specific and nonchronic pain), and health conditions (hypertension, angina, allergies, and trouble sleeping).

The inverse probability weighted samples of those completing wave 2 only and those completing waves 2 and 3 were similar demographically to the sample of all individuals who completed wave 1 for all categories except for income (Table S2 in Multimedia Appendix 1) [49]. Income was similar for the unweighted data and sample weights produced differences in the estimate of some income categories like the US $30,000-US $39,999 income range [50]. Estimated sample weights and weighted sample distributions did not vary between the backward elimination estimation–derived weights and those not using backward elimination estimation and were robust to model specification.

The greatest differences observed between the baseline and the weighted samples were in income, where the weighted sample had approximately US $5000 more in annual income than the baseline population. In addition, the wave 2 and 3 weighted samples had lower measured scores on the ISS, the ODI, the RMDQ, the PEG, and the SBST compared with the weighted participants in waves 2 only and the unweighted baseline sample (Table S3 and S4 in Multimedia Appendix 1). These differences were less than those between the unweighted samples (Table 4).

Respondents were generally older, less Hispanic, less likely to have a bachelor's degree, less likely to be married, lower income, and less likely to be employed full-time when comparing unweighted and weighted wave 2 responses (Table S2 in Multimedia Appendix 1). Unweighted respondents had fewer health conditions and were less likely to have all health conditions except arthritis, anxiety, allergies, and trouble sleeping than weighted respondents in Wave 2 (Table S3 and S4 in Multimedia Appendix 1). The largest difference between unweighted and weighted respondents in wave 2 was in pain assessments. Unweighted respondents were much less likely to have nonspecific pain (0.55 vs 0.74), more likely to have chronic pain (0.92 vs 0.88) and have lower pain intensity and pain interference than weighted respondents. Unweighted respondents also scored systematically lower on the ISS, the ODI, the RMDQ, the PEG, and the SBST.

## Discussion

### Principal Results and Comparison With Previous Work

Comparisons of MTurk samples with the general US population have shown that MTurk participants tend to be younger, more educated, and less racially and ethnically diverse than the general population [11]. The characteristics of our baseline sample are consistent with this literature. This study extends the existing literature by describing nonresponse rates and predictors of nonresponse in a longitudinal study of MTurk participants with back pain, a population increasingly studied

using survey methods to assess severity and changes over time [30,31].

We found that overall nonresponse was larger from wave 1 to wave 2 than between waves 2 and 3, even with a sample that was screened for higher quality respondents and those with lower quality responses were removed; however, it should be noted that since participants already responded in wave 2, there is selection bias as they have already participated in a follow-up survey wave. Previous research has shown that the majority of MTurk workers are frequent users of MTurk [51], participating in many tasks and using the platform regularly. Despite this, we found that even with regular outreach about the follow-up survey, response rates declined. While our survey focused on patients with back pain, our retention rate is on par with longitudinal studies that included more general samples of MTurk respondents [52], and greater than the average internet survey response rate of 40%-50% [53].

In addition, we found differential response rates by participant characteristics. Response rates were generally over 40% for all categories except those with relatively small sample sizes (ie, <2% of the overall sample). In this study, we have sample coverage for subgroups that reflect the larger population, indicating higher "positivity" as per Mercer et al [27]. Persistent respondents in our survey were older with higher income, less likely to be Hispanic, less educated, less likely to be employed full-time, and more likely to never have been married, all broadly consistent with previous nonresponse studies in MTurk [19]. In comparing response rates from wave 2 to wave 3, we also found less differential nonresponse than when comparing wave 2 with wave 1 and waves 2 and 3 with wave 1.

Participants who continued to respond in survey waves completed their previous surveys faster than those who did not respond and reported fewer health conditions. Our multivariate models largely supported these findings. Factors associated with increased response across survey waves were completing our previous survey faster, being divorced or never marrying, and having allergies or trouble sleeping, while factors negatively associated with response were being Hispanic or Latino, younger, having a bachelor's degree, and having more reported pain intensity and interference. Our backward estimation specification allowed for individual categorical responses (ie, income, education, and employment) to be assessed independently, leading to instances where only specific subsets of a category are included in the model. The reference groups in these cases are any individuals not included in that category, sometimes making model interpretation unintuitive (ie, having an income of US $40,000-US $59,999). However, categories with the lowest response rates compared with wave 1 were included in the models, further supporting the model for the prediction of nonresponse. The effect of the inverse probability weights was robust to whether the subset of significant variables or all candidate variables were used to derive them. All of this mitigated "composition" issues that limit the ability to draw inferences as per Mercer et al [27].

The inverse probability weighted sample analysis produced sample demographics and condition prevalence rates that were similar between baseline respondents and those responding at subsequent waves except for pain assessments, which is most likely driven by changes in pain assessments over time rather than bias introduced by nonresponse. Average pain, according to different measures, was higher in the weighted sample than the unweighted samples, indicating that those with higher pain are more likely to be due to dropout and less likely to be due to a reduction in pain over time. Given the weighting was able to recreate wave 1 characteristic distributions even though the wave 2 and waves 2 and 3 samples were not similar to wave 1, we ensure that the composition of the weighted sample reflects our starting sample and would minimize inferential bias. These results were consistent with descriptive analysis of the survey waves, highlighting a potential issue with longitudinal analyses on those who report pain.

In each successive wave, those with more, nonchronic, and specific pain had a higher likelihood of nonresponse, potentially biasing analyses that do not account for differences in panel composition over time. Analyses using convenience panels need to account for changes in sample composition, particularly the loss of those with more severe and specific pain, and worse health. Estimates of impacts on pain may be biased if those with the most pain and the more specific pain are differentially dropping out of the panel. Hence, evaluating whether attrition impacts inferences about change in outcome measures such as pain is important.

Platforms like MTurk will continue to be used to collect survey data inexpensively and rapidly. Results from this nonresponse analysis should be considered in the context of general trends for MTurk workers. First, researchers should consider the underlying population that completes surveys on MTurk. Generally, the population who responds is younger, so higher nonresponse rates among younger individuals may be less concerning as that subgroup is already providing a higher proportion of responses than the general US population [13]. In addition, those with back pain tend to be older on average and come from predominantly non-White racial and ethnic groups [28]. However, high nonresponse rates among Hispanic or Latino participants may bias inferences as they represent smaller proportions of the base sample. In addition, analyses that account for specific populations like those with back pain should account for differences in sample composition over time, particularly around variables that may impact response rates and potential outcomes such as pain intensity or severity.

Regular outreach and follow-up have already been shown to increase response rates in longitudinal studies with MTurk. Inverse probability weighting can be used to correct for participant nonresponse. Given patterns of nonresponse, targeted surveys to populations that are either underrepresented or are more likely to drop out of surveys with MTurk can avoid issues with weighting small samples, especially when multiple successive survey waves exist.

## Limitations

This study had several limitations. First, since it focused on patients with back pain, its results should be interpreted with care and may not generalize to those without back pain. In addition, all data collection occurred during the COVID-19 pandemic. Response behavior may have differed during this

time compared with before or after the pandemic. Finally, all data related to health conditions and pain assessments were self-reported. However, we applied various approaches to limit our sample to high-quality and truthful responses.

## Conclusions

Longitudinal studies on MTurk, particularly those exploring specific issues like the impacts of pain on health, should carefully consider how nonresponse will affect their study samples and whether samples drawn from these studies reflect the population of interest.

## Acknowledgments

The authors disclosed the following financial support for the research, authorship, and publication of this article: This work was supported by the National Center for Complementary and Integrative Health (NCCIH; grant 1R01AT010402-01A1). NCCIH had no role in the design, data collection, analysis, interpretation, or writing of this manuscript.

## Data Availability

The dataset analyzed for the current study is not publicly available yet due to the project still being in progress, but the data are available from the first author upon reasonable request.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist and additional tables (weighted responses).
[DOCX File , 34 KB - ijmr_v13i1e58771_app1.docx ]

## References

1. Chandler J, Shapiro D. Conducting clinical research using crowdsourced convenience samples. Annu Rev Clin Psychol 2016;12:53-81. [doi: 10.1146/annurev-clinpsy-021815-093623] [Medline: 26772208]
2. Gosling SD, Mason W. Internet research in psychology. Annu Rev Psychol 2015 Jan 03;66:877-902. [doi: 10.1146/annurev-psych-010814-015321] [Medline: 25251483]
3. Mortensen K, Hughes TL. Comparing Amazon's Mechanical Turk platform to conventional data collection methods in the health and medical research literature. J Gen Intern Med 2018 Apr;33(4):533-538 [FREE Full text] [doi: 10.1007/s11606-017-4246-0] [Medline: 29302882]
4. Aguinis H, Villamor I, Ramani RS. MTurk research: review and recommendations. Journal of Management 2020 Nov 19;47(4):823-837. [doi: 10.1177/0149206320969787]
5. Zhang B, Gearhart S. Collecting Online Survey Data: A Comparison of Data Quality among a Commercial Panel & MTurk. Surv Pract 2020;13(1):1-10. [doi: 10.29115/sp-2020-0015]
6. Stoycheff E. Please participate in Part 2: Maximizing response rates in longitudinal MTurk designs. Methodological Innovations 2016 Oct 12;9:205979911667287. [doi: 10.1177/2059799116672879]
7. Callegaro, M, Villar A, Yeager D, Krosnick J. A Critical Review of Studies Investigating The Quality of Data Obtained With Online Panels Based On Probability And Nonprobability Samples1. John Wiley UK: 11 April2014; 2014:53.
8. Lehdonvirta V, Oksanen A, Räsänen P, Blank G. Social media, web, and panel surveys: using non‐probability samples in social and policy research. Policy & Internet 2020 Apr 29;13(1):134-155. [doi: 10.1002/poi3.238]
9. Hilton LG, Coulter ID, Ryan GW, Hays RD. Comparing the recruitment of research participants with chronic low back pain using Amazon Mechanical Turk with the recruitment of patients from chiropractic clinics: a quasi-experimental study. J Manipulative Physiol Ther 2021 Oct;44(8):601-611 [FREE Full text] [doi: 10.1016/j.jmpt.2022.02.004] [Medline: 35728997]
10. Levay KE, Freese J, Druckman JN. The demographic and political composition of Mechanical Turk samples. Sage Open 2016 Mar 15;6(1):215824401663643. [doi: 10.1177/2158244016636433]
11. Mortensen K, Alcalá MG, French MT, Hu T. Self-reported health status differs for Amazon's Mechanical Turk respondents compared with nationally representative surveys. Med Care 2018 Mar;56(3):211-215. [doi: 10.1097/MLR.0000000000000871] [Medline: 29329148]
12. Moss A. Demographics of People on Amazon Mechanical Turk. 2020. URL: https://www.cloudresearch.com/resources/blog/who-uses-amazon-mturk-2020-demographics/ [accessed 2024-03-09]
13. Qureshi N, Edelen MO, Hilton L, Rodriguez A, Hays RD, Herman PM. Comparing Data Collected on Amazon's Mechanical Turk to National Surveys. Am J Health Behav 2022 Oct 17;46(5):497-502. [doi: 10.5993/AJHB.46.5.1] [Medline: 36333833]

14. Difallah D, Filatova E, Ipeirotis P. Demographics and dynamics of mechanical turk workers. 2018 Presented at: Proceedings of the 11th ACM International Conference on Web Search and Data Mining; 2018Feb 2; Marina Del Rey ,US p. 135-143. [doi: 10.1145/3159652.3159661]
15. Cornesse C, Annelies GB, Dutwin, D, Krosnick, JA, De Leeuw, ED, Legleye S, et al. A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. Journal of Survey Statistics and Methodology 2020;8(1):4-36. [doi: 10.1093/jssam/smz041]
16. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: a meta-analysis. Public Opinion Quarterly 2008;72(2):167-189. [doi: 10.1093/poq/nfn011]
17. Zack, ES, Kennedy J, Long JS. Can nonprobability samples be used for social science research? a cautionary tale. Surv. Res. Methods 2019;13(2):215-227. [doi: 10.18148/srm/2019.v13i2.7262]
18. Zhou H, Fishbach A. The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. J Pers Soc Psychol 2016 Oct;111(4):493-504. [doi: 10.1037/pspa0000056] [Medline: 27295328]
19. Yu T, Chen J, Gu NY, Hay JW, Gong CL. Predicting panel attrition in longitudinal HRQoL surveys during the COVID-19 pandemic in the US. Health Qual Life Outcomes 2022 Jul 06;20(1):104 [FREE Full text] [doi: 10.1186/s12955-022-02015-8] [Medline: 35794553]
20. Rung JM, Almog S, Ferreiro AV, Berry MS. Using crowdsourcing for alcohol and nicotine use research: prevalence, data quality, and attrition on Amazon Mechanical Turk. Subst Use Misuse 2022;57(6):857-866 [FREE Full text] [doi: 10.1080/10826084.2022.2046096] [Medline: 35258409]
21. Banks J, Muriel A, Smith JP. Attrition and health in ageing studies: Evidence from ELSA and HRS. Longit Life Course Stud 2011;2(2):10 [FREE Full text] [doi: 10.14301/llcs.v2i2.115] [Medline: 24376472]
22. McDonald B, Haardoerfer R, Windle M, Goodman M, Berg C. Implications of attrition in a longitudinal web-based survey: an examination of college students participating in a tobacco use study. JMIR Public Health Surveill 2017 Oct 16;3(4):e73 [FREE Full text] [doi: 10.2196/publichealth.7424] [Medline: 29038092]
23. Rübsamen N, Akmatov MK, Castell S, Karch A, Mikolajczyk RT. Factors associated with attrition in a longitudinal online study: results from the HaBIDS panel. BMC Med Res Methodol 2017 Aug 31;17(1):132 [FREE Full text] [doi: 10.1186/s12874-017-0408-3] [Medline: 28859617]
24. Minderop I, Weiß B. Now, later, or never? Using response-time patterns to predict panel attrition. International Journal of Social Research Methodology 2022 Jul 08;26(6):693-706. [doi: 10.1080/13645579.2022.2091259]
25. Chatfield MD, Brayne CE, Matthews FE. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. J Clin Epidemiol 2005 Jan;58(1):13-19. [doi: 10.1016/j.jclinepi.2004.05.006] [Medline: 15649666]
26. Gustavson K, Røysamb E, Borren I. Preventing bias from selective non-response in population-based survey studies: findings from a Monte Carlo simulation study. BMC Med Res Methodol 2019 Jun 13;19(1):120 [FREE Full text] [doi: 10.1186/s12874-019-0757-1] [Medline: 31195998]
27. Mercer AW, Kreuter F, Keeter S, Stuart EA. Theory and practice in nonprobability. Parallels Between Causal Inference And Survey Inference 2017:250-271. [doi: 10.1093/poq/nfw060]
28. Lucas JW, Connor EM, Bose J. Back, lower limb, and upper limb pain among U.S. adults, 2019. NCHS Data Brief 2021 Jul(415):1-8 [FREE Full text] [Medline: 34473621]
29. Murray CJL, Atkinson C, Bhalla K, Birbeck G, Burstein R, Chou D, Foreman, Lopez, Murray, Dahodwala, Jarlais, Fahami, Murray, Jarlais, Foreman, Lopez, Murray, U.S. Burden of Disease Collaborators. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. JAMA 2013 Aug 14;310(6):591-608 [FREE Full text] [doi: 10.1001/jama.2013.13805] [Medline: 23842577]
30. Keyes KM, Rutherford C, Popham F, Martins SS, Gray L. How healthy are survey respondents compared with the general population?: Using survey-linked death records to compare mortality outcomes. Epidemiology 2018 Mar;29(2):299-307 [FREE Full text] [doi: 10.1097/EDE.0000000000000775] [Medline: 29389712]
31. Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A demonstration of the impact of response bias on the results of patient satisfaction surveys. Health Serv Res 2002 Oct;37(5):1403-1417 [FREE Full text] [doi: 10.1111/1475-6773.11194] [Medline: 12479503]
32. Stewart N, Ungemach C, Harris AJL, Bartels DM, Newell BR, Paolacci G, et al. The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. Judgm. decis. mak 2023 Jan 01;10(5):479-491. [doi: 10.1017/s1930297500005611]
33. CloudResearch. CloudResearch | Online Research & Participant Recruitment Made Easy. 2023. URL: https://www.cloudresearch.com/ [accessed 2023-02-28]
34. Peer E, Vosgerau J, Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. Behav Res Methods 2014 Dec;46(4):1023-1031. [doi: 10.3758/s13428-013-0434-y] [Medline: 24356996]
35. Robinson J, Rosenzweig C, Moss AJ, Litman L. Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. PLoS One 2019;14(12):e0226394 [FREE Full text] [doi: 10.1371/journal.pone.0226394] [Medline: 31841534]

36. Hauser DJ, Moss AJ, Rosenzweig C, Jaffe SN, Robinson J, Litman L. Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. Behav Res Methods 2023 Dec;55(8):3953-3964 [FREE Full text] [doi: 10.3758/s13428-022-01999-x] [Medline: 36326997]
37. Herman P, Hays R. Open ICPSR. Measuring Chronic Pain Impact. URL: https://www.openicpsr.org/openicpsr/project/198049/version/V1/view
38. Hays RD, Qureshi N, Herman PM, Rodriguez A, Kapteyn A, Edelen MO. Effects of excluding those who report having "Syndomitis" or "Chekalism" on data quality: longitudinal health survey of a sample from Amazon's Mechanical Turk. J Med Internet Res 2023 Aug 04;25:e46421 [FREE Full text] [doi: 10.2196/46421] [Medline: 37540543]
39. Qureshi N, Hays RD, Herman PM. Are some ways of defining chronic low back pain more indicative of future back pain than others? Pain Med 2024 Feb 01;25(2):160-162. [doi: 10.1093/pm/pnad135] [Medline: 37773991]
40. Hays RD, Edelen MO, Rodriguez A, Herman P. Support for the reliability and validity of the National Institutes of Health Impact Stratification Score in a sample of active-duty U.S. Military personnel with low back pain. Pain Med 2021 Oct 08;22(10):2185-2190 [FREE Full text] [doi: 10.1093/pm/pnab175] [Medline: 34022052]
41. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000 Nov 15;25(22):2940-52; discussion 2952. [doi: 10.1097/00007632-200011150-00017] [Medline: 11074683]
42. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. Phys Ther 2016 Oct;96(10):1620-1637. [doi: 10.2522/ptj.20150420] [Medline: 27081203]
43. Krebs EE, Lorenz KA, Bair MJ, Damush TM, Wu J, Sutherland JM, et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. J Gen Intern Med 2009 Jun;24(6):733-738 [FREE Full text] [doi: 10.1007/s11606-009-0981-1] [Medline: 19418100]
44. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 2008 May 15;59(5):632-641 [FREE Full text] [doi: 10.1002/art.23563] [Medline: 18438893]
45. Heinze G, Dunkler D. Five myths about variable selection. Transpl Int 2017 Jan;30(1):6-10 [FREE Full text] [doi: 10.1111/tri.12895] [Medline: 27896874]
46. Smith G. Step away from stepwise. J Big Data 2018 Sep 15;5(1):32. [doi: 10.1186/s40537-018-0143-6]
47. Schmidt CO, Raspe H, Pfingsten M, Hasenbring M, Basler HD, Eich W, et al. Does attrition bias longitudinal population-based studies on back pain? Eur J Pain 2011 Jan;15(1):84-91. [doi: 10.1016/j.ejpain.2010.05.007] [Medline: 20542714]
48. StataCorp. Stata Statistical Software: Release 17. College Station, TX. 2024. URL: https://www.stata.com/support/faqs/resources/citing-software-documentation-faqs/ [accessed 2024-10-09]
49. Bu F. Non-response and attrition in longitudinal studies. J Epidemiol Community Health 2022 Oct 28. [doi: 10.1136/jech-2022-219861] [Medline: 36307201]
50. Haddad C, Sacre H, Zeenny RM, Hajj A, Akel M, Iskandar K, et al. Should samples be weighted to decrease selection bias in online surveys during the COVID-19 pandemic? Data from seven datasets. BMC Med Res Methodol 2022 Mar 06;22(1):63 [FREE Full text] [doi: 10.1186/s12874-022-01547-3] [Medline: 35249541]
51. Pew Research Center. Research In The Crowdsourcing Age, A Case Study. 2024. URL: https://www.pewresearch.org/internet/2016/07/11/turkers-in-this-canvassing-young-well-educated-and-frequent-users/ [accessed 2024-03-13]
52. Shapiro DN, Chandler J, Mueller PA. Using Mechanical Turk to study clinical populations. Clinical Psychological Science 2013 Jan 31;1(2):213-220. [doi: 10.1177/2167702612469015]
53. Wu MJ, Zhao K, Fils-Aime F. Response rates of online surveys in published research: A meta-analysis. Computers in Human Behavior Reports 2022 Aug;7:100206. [doi: 10.1016/j.chbr.2022.100206]


## Abbreviations

**HIT:** human intelligence task
**ISS:** Impact Stratification Score
**MTurk:** Amazon Mechanical Turk
**ODI:** Oswestry Disability Index
**OR:** odds ratio
**PEG:** Pain, Enjoyment of Life and General Activity scale
**RMDQ:** Roland Morris Disability Questionnaire
**SBST:** Keele STarT Back Screening Tool
**STROBE:** Strengthening the Reporting of Observational Studies in Epidemiology

*Edited by T de Azevedo Cardoso; submitted 25.03.24; peer-reviewed by T Yu, AJ Moss, E Crespi; comments to author 07.06.24; revised version received 21.06.24; accepted 25.09.24; published 11.11.24.*

*Please cite as:*
*Qureshi N, Hays RD, Herman PM*
*Dropout in a Longitudinal Survey of Amazon Mechanical Turk Workers With Low Back Pain: Observational Study*
*Interact J Med Res 2024;13:e58771*
*URL: https://www.i-jmr.org/2024/1/e58771*
*doi:10.2196/58771*
*PMID:*

Original Paper

# Evaluation of the Accuracy, Credibility, and Readability of Statin-Related Websites: Cross-Sectional Study

Eunice Ling[1], BAppSc, MD; Domenico de Pieri[1*], MD; Evenne Loh[1*], MD; Karen M Scott[1], BEd, MA, PhD; Stephen C H Li[2], MBBS, MSci, MPH, MBA, MAACB; Heather J Medbury[1,3], BSc, PhD

[1]Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[2]Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, Australia

[3]Vascular 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.

***(Interact J Med Res 2024;13:e42849)*** doi:10.2196/42849

**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]. Health literacy–related knowledge and skills are particularly deficient among vulnerable populations, who are also more likely to experience CVD [12,13]. Unfortunately, those with poor health literacy are susceptible to the influence of mass media and emotionally persuasive texts. This may explain the response to the television program "Catalyst" in Australia [14], whose criticism of statins resulted in 11% of patients (in a survey by the Australian National Heart Foundation of 1094 patients) who watched the program ceasing to take their cholesterol-lowering medication and significant and sustained changes in statin usage, with 2.6% fewer statins (equivalent to 14,005 dispensing) each week [15].

## Internet Standards

There is considerable heterogeneity in the quality of web-based health information [16]. The quality of information can be examined in 3 domains: accuracy, credibility, and readability [17]. Each is defined as follows: accuracy is the intent to be evidence based and safe by adequately offering a complete, unbiased picture and its relevance [17]; credibility is the attribution of source and authorship and the disclosure of conflicts of interest for the presented information [18]; and readability is the ease of understanding due to the style of writing, describing the reading comprehension level a layperson requires to understand a text [19,20]. Complex wording reduces engagement with, and application of, content [21]. This leaves patients vulnerable to becoming ill-informed and at risk of adverse health outcomes [22]. The quality of web-based health information is, therefore, a pressing issue, which the Health on the Net Foundation aims to address by providing a Health on the Net Code of Conduct (HONcode) seal, an internet-based certification of medical and health websites that adheres to a set of publishing principles regarding the source and purpose of medical information. However, few consumers are aware of Health on the Net or the HONcode.

While there have been assessments of the quality of websites on many key areas of health care, such as diabetes [23], obesity [24], and hypertension [25], and on surgical interventions [26,27], there is a dearth of assessment about the quality of web-based information pertaining to medications. Such assessments are necessary to inform clinicians of the quality of content likely to be accessed by their patients, who are particularly interested in the likelihood and nature of adverse side effects. We aimed to assess the quality of consumer health information on websites about statins through consideration of accuracy, credibility, and readability.

# Methods

## Study Design

### *Overview*

Through this cross-sectional study, we analyzed the accuracy, credibility, and readability of websites that were most commonly presented to patients searching for the keywords "statin" and "statin side effects." We determined each website's search engine ranking, category (commercial, not-for-profit, or media), and the presence or absence of the HONcode seal. Furthermore,

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 results 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.

## Measures

### *Overview*

In assessing the accuracy, credibility, and readability of the websites, we considered existing tools and developed study-specific tools where necessary.

### *Accuracy*

For accuracy, we developed a statin-specific tool that took into account medical guidelines. The checklist (Table 1) was designed after referring to other studies purporting to assess the quality of web-based health information [17]. A key difference here was that we were looking at a specific treatment. Three features were considered: (1) the intent to be evidence based; (2) safety, in that a website should adequately offer users a complete, unbiased picture of statin treatment; and (3) relevance, in that it is reasonable to expect a website to address the criteria in the checklist [17]. The checklist was intrinsically linked to a website's comprehensiveness, consistent with other studies that have evaluated completeness as an integral part of accuracy [16].

To meet these features, guidelines from the American College of Cardiology and the Australian Heart Foundation [37] were synthesized into short statements, which formed the accuracy criteria. These statements formed a checklist that each website was required to address to be considered "accurate." The development of the statements was further informed and determined by a review of the treatment of cholesterol in light of its evidence base [38], as well as criteria from the treatment section of the validated DISCERN tool [39]. This section (items 9-16 of the DISCERN tool) addresses issues of risk, benefit, and how the website guides decision-making surrounding treatment options [39]. Combining these sources ensured that a higher score would be awarded to websites providing the most evidence-based information. In total, 20 equally weighted criteria were devised, and a score of 20 was arbitrarily defined as a minimum acceptable standard.

Each accuracy item was scored as: "present and complete" (2), "present but incomplete" (1), "absent" (0), or "inaccurate" (*). A maximum score of 40 could be awarded to each website assessed. "Incompleteness" was defined as a nonexact or indirect mention of a topic outlined in a criterion rather than an explicit statement. Two reviewers (DdP and E loh) completed the assessment.

Any hyperlinks that navigated to information within a website were followed and the data were included in the final assessment; links leading to external websites were not followed. Embedded videos were analyzed. Once each reviewer concluded their analysis, the results were compared. Discrepancies were resolved through discussion until reaching a consensus.

**Table 1.** Website accuracy checklist.

| Accuracy criteria | Descriptor |
|---|---|
| 1 | Mentions that cholesterol is a modifiable risk factor for cardiovascular disease [35] |
| 2 | Mentions that consultation with a doctor is essential before and while taking statins and when ceasing them [36] |
| 3 | Lists conditions for which statins are used [36] |
| 4 | Defines the target population for statin therapy [37] |
| 5 | Mentions the importance of adherence to statin therapy [36] |
| 6 | Addresses the subtleties of primary prevention [38] |
| 7 | Mentions that statins are about reducing complications of high cholesterol rather than achieving a specific (low-density lipoprotein) cholesterol [38] |
| 8 | Describes or at least lists the benefits of statin therapy [39] |
| 9 | Describes or at least lists the side effects or risks of statin therapy [39] |
| 10 | Describes how treatment affects the overall quality of life [39] |
| 11 | Mentions low to moderate dose statin therapy is recommended in primary prevention [37] |
| 12 | Specifically addresses rhabdomyolysis [36] |
| 13 | Describes the approximate financial burden to the patient [36] |
| 14 | Describes the duration of treatment before an effect is measurable [39] |
| 15 | Describes how statins work or at least what they do [39] |
| 16 | Describes what may happen without treatment [39] |
| 17 | Explores the possibility of using alternative therapies to statins [39] |
| 18 | Mentions that statins must not be used during pregnancy [37] |
| 19 | Describes drug interactions or at least lists them [39] |
| 20 | Mentions that statins do not replace a healthy lifestyle [36] |

### *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 2.** Website credibility checklist.

| Item | Criteria | Score criteria |
|---|---|---|
| 1 | Referencing or citations obtained from peer-reviewed journals | 1 point if the articles for which the references are obtained are published in peer-reviewed journals [16,44] |
| 2 | Website updated within last 24 months | The latest update should be within the past 24 months [45] |
| 3 | Avoids anecdotal evidence for making claims | Does not use anecdotal evidence as a basis for claims; quoting a case study without using claims is acceptable [46] |
| 4 | Mailing address present | Physical contact address of the website clearly stated [45] |
| 5 | Contact information available | Contact information including name, position, telephone number, address, and email [47] |
| 6 | Sponsorship stated | Any sponsorship should be clearly stated |
| 7 | Organizational privacy policy stated | Organization privacy policy should be clearly stated [47] |
| 8 | Declaration of the author's qualification | Author's qualification should be health care related [16,44] |
| 9 | Paid access tab present | If paid access is available, the difference in the information obtained from paid vs unpaid access should be clearly stated [48] |
| 10 | Disclosure of funding or conflicts of interest | Conflicts of interest and funding disclosure should be clearly stated [44] |
| 11 | The presence of an HONcode seal or third-party certification | Presence of a HONcode[a] seal or any other third-party certification [16] |
| 12 | Advertisement neutral | Advertisements should steer clear from the website information (eg, no pop-ups related to the website content) [16] |
| 13 | Disclaimer regarding web-based health information | A disclaimer should be clearly stated that web-based health information does not replace a practitioner's advice [44] |

[a]HONcode: 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[a] presence or absence, accuracy, credibility, and readability.

| Website rank | Category | HONcode Seal | Accuracy | Credibility | Readability (SMOG[b] score) |
|---|---|---|---|---|---|
| 1 | Commercial | Yes | 20 | 6 | 10.7 |
| 2 | Not-for-profit | Yes | 9 | 11 | 11.3 |
| 3 | Not-for-profit | No | 23 | 9 | 12.5 |
| 4 | Commercial | Yes | 16 | 6 | 13 |
| 5 | Commercial | No | 10 | 4 | 7.7 |
| 6 | Not-for-profit | No | 15 | 7 | 13 |
| 7 | Commercial | No | 14 | 10 | 9.6 |
| 8 | Not-for-profit | No | 13 | 5 | 10.8 |
| 9 | Not-for-profit | No | 16 | 11 | 10.2 |
| 10 | Commercial | Yes | 11 | 10 | 8.6 |
| 11 | Commercial | Yes | 16 | 9 | 9.4 |
| 12 | Commercial | No | 13 | 6 | 11 |
| 13 | Not-for-profit | No | 17 | 9 | 8.9 |
| 14 | Not-for-profit | No | 8 | 7 | 9 |
| 15 | Not-for-profit | No | 8 | 7 | 10.3 |
| 16 | Media | No | 8 | 4 | 7.8 |
| 17 | Not-for-profit | Yes | 16 | 11 | 6.7 |
| 18 | Commercial | Yes | 19 | 8 | 9.7 |
| 19 | Commercial | Yes | 21 | 7 | 9.2 |
| 20 | Media | No | 14 | 2 | 8.1 |

[a]HONcode: Health on the Net Code of Conduct.

[b]SMOG: Simple Measure of Gobbledegook.

## 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.

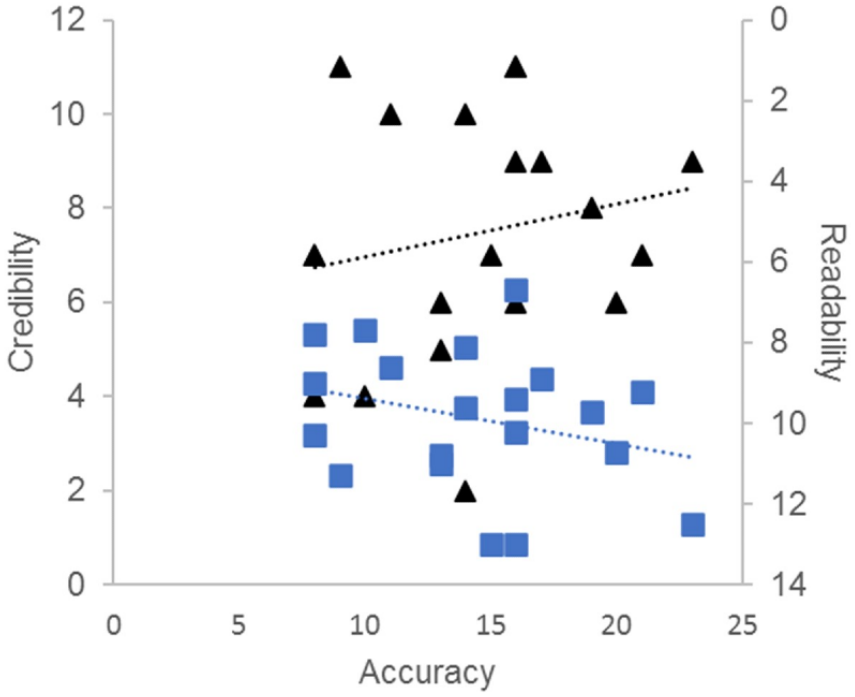
## Readability

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 credibilty 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

[12,13], making them especially vulnerable to inaccurate, untrustworthy, and unreadable websites on statins.

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.
[DOCX File , 14 KB - ijmr_v13i1e42849_app1.docx ]

## References

1. Cardiovascular diseases (CVDs). World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2023-12-29]
2. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. Lancet 2016;388(10059):2532-2561 [FREE Full text] [doi: 10.1016/S0140-6736(16)31357-5] [Medline: 27616593]
3. Grad R, Légaré F, Bell NR, Dickinson JA, Singh H, Moore AE, et al. Shared decision making in preventive health care: what it is; what it is not. Can Fam Physician 2017;63(9):682-684 [FREE Full text] [Medline: 28904031]
4. Mozafarpour S, Norris B, Borin J, Eisner BH. Assessment of readability, quality and popularity of online information on ureteral stents. World J Urol 2018;36(6):985-992 [FREE Full text] [doi: 10.1007/s00345-018-2179-9] [Medline: 29435639]
5. Krot K. Online information seeking and trust in doctors: an empirical study. 2017 Presented at: 24th International Scientific Conference on Economic and Social Development – "Managerial Issues in Modern Business"; October 13-14, 2017; Warsaw.
6. Kim H, Park SY, Bozeman I. Online health information search and evaluation: observations and semi-structured interviews with college students and maternal health experts. Health Info Libr J 2011;28(3):188-199 [FREE Full text] [doi: 10.1111/j.1471-1842.2011.00948.x] [Medline: 21831218]
7. Wosik J, Clowse MEB, Overton R, Adagarla B, Economou-Zavlanos N, Cavalier J, et al. Impact of the COVID-19 pandemic on patterns of outpatient cardiovascular care. Am Heart J 2021;231:1-5 [FREE Full text] [doi: 10.1016/j.ahj.2020.10.074] [Medline: 33137309]
8. Anderson J, Rainie L, Vogels EA. Experts say the 'new normal' in 2025 will be far more tech-driven, presenting more big challenges. Pew Research Center.: Elon University; 2021. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/02/PI_2021.02.18_New-Normal-2025_FINAL.pdf [accessed 2023-12-29]
9. Institute of Medicine, Board on Neuroscience and Behavioral Health, Committee on Health Literacy. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health Literacy: A Prescription to End Confusion. Washington (DC): National Academies Press; 2004.
10. Peerson A, Saunders M. Health literacy revisited: what do we mean and why does it matter? Health Promot Int 2009;24(3):285-296 [FREE Full text] [doi: 10.1093/heapro/dap014] [Medline: 19372101]
11. Mackert M, Kahlor L, Tyler D, Gustafson J. Designing e-health interventions for low-health-literate culturally diverse parents: addressing the obesity epidemic. Telemed J E Health 2009;15(7):672-677 [FREE Full text] [doi: 10.1089/tmj.2009.0012] [Medline: 19694596]
12. Keselman A, Smith CA, Murcko AC, Kaufman DR. Evaluating the quality of health information in a changing digital ecosystem. J Med Internet Res 2019;21(2):e11129 [FREE Full text] [doi: 10.2196/11129] [Medline: 30735144]
13. Havranek EP, Mujahid MS, Barr DA, Blair IV, Cohen MS, Cruz-Flores S, et al. Social determinants of risk and outcomes for cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2015;132(9):873-898 [FREE Full text] [doi: 10.1161/CIR.0000000000000228] [Medline: 26240271]
14. Demasi M, Arnott I, editors. Heart of the Matter. Australia: Australian Broadcasting Corporation; 2013.
15. Schaffer AL, Buckley NA, Dobbins TA, Banks E, Pearson SA. The crux of the matter: did the ABC's catalyst program change statin use in Australia? Med J Aust 2015;202(11):591-595 [FREE Full text] [doi: 10.5694/mja15.00103] [Medline: 26068693]

16. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700 [FREE Full text] [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
17. Ahmed OH, Sullivan SJ, Schneiders AG, McCrory PR. Concussion information online: evaluation of information quality, content and readability of concussion-related websites. Br J Sports Med 2012;46(9):675-683. [doi: 10.1136/bjsm.2010.081620] [Medline: 21504964]
18. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
19. Klare GR. The measurement of readability: useful information for communicators. ACM J Comput Doc 2000;24(3):107-121 [FREE Full text] [doi: 10.1145/344599.344630]
20. Alamoudi U, Hong P. Readability and quality assessment of websites related to microtia and aural atresia. Int J Pediatr Otorhinolaryngol 2015;79(2):151-156 [FREE Full text] [doi: 10.1016/j.ijporl.2014.11.027] [Medline: 25554575]
21. Australia's Health 2018: In Brief. Canberra: Australian Institute of Health and Welfare; 2018.
22. McInnes N, Haglund BJA. Readability of online health information: implications for health literacy. Inform Health Soc Care 2011;36(4):173-189 [FREE Full text] [doi: 10.3109/17538157.2010.542529] [Medline: 21332302]
23. Lipari M, Berlie H, Saleh Y, Hang P, Moser L. Understandability, actionability, and readability of online patient education materials about diabetes mellitus. Am J Health Syst Pharm 2019;76(3):182-186 [FREE Full text] [doi: 10.1093/ajhp/zxy021] [Medline: 31408087]
24. El-Haddad N, Spooner C, Faruqi N, Denney-Wilson E, Harris M. Readability and content analysis of lifestyle education resources for weight management in Australian general practice. BMC Obes 2016;3:16 [FREE Full text] [doi: 10.1186/s40608-016-0097-1] [Medline: 26966546]
25. Gulati R, Nawaz M, Lam L, Pyrsopoulos NT. Comparative readability analysis of online patient education resources on inflammatory bowel diseases. Can J Gastroenterol Hepatol 2017;2017:3681989 [FREE Full text] [doi: 10.1155/2017/3681989] [Medline: 28740843]
26. Grewal P, Williams B, Alagaratnam S, Neffendorf J, Soobrah R. Quality of vascular surgery web sites on the internet. J Vasc Surg 2012;56(5):1461-1467 [FREE Full text] [doi: 10.1016/j.jvs.2012.04.058] [Medline: 22801107]
27. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. Clin Orthop Relat Res 2010;468(10):2572-2580 [FREE Full text] [doi: 10.1007/s11999-010-1380-y] [Medline: 20496023]
28. Browser market share Australia. Statcounter. 2021. URL: https://gs.statcounter.com/browser-market-share/all/australia [accessed 2021-09-04]
29. Martin J. Best web browsers. Tech Advisor. 2021. URL: https://www.techadvisor.com/article/728377/best-web-browsers-2021.html [accessed 2021-09-04]
30. Baker J. Brafton 2017 content marketing benchmark report. Brafton. 2017. URL: https://www.brafton.com/blog/strategy/brafton-2017-content-marketing-benchmark-report/ [accessed 2021-09-04]
31. Chitika insights: the value of Google result positioning. Chitika. 2013. URL: https://research.chitika.com/wp-content/uploads/2022/02/chitikainsights-valueofgoogleresultspositioning.pdf [accessed 2023-12-29]
32. Purcell K, Brenner J, Rainie L. Search engine use 2012. Pew Research Centre. 2012. URL: https://www.ris.org/uploadi/editor/1341041853PIP_Search_Engine_Use_2012.pdf [accessed 2023-12-29]
33. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004;7(5):497-510. [doi: 10.1089/cpb.2004.7.497] [Medline: 15667044]
34. Google organic CTR history. Advanced Web Ranking. 2022. URL: https://www.advancedwebranking.com/ctrstudy/ [accessed 2023-12-29]
35. Wilson PWF. Overview of established risk factors for cardiovascular disease. UpToDate. 2021. URL: https://www.uptodate.com/contents/overview-of-established-risk-factors-for-cardiovascular-disease#topicContent [accessed 2021-09-04]
36. Davidson P, Chew D. Reducing Risk in Heart Disease: An Expert Guide to Clinical Practice for Secondary Prevention of Coronary Heart Disease. Australia: National Heart Foundation of Australia; 2012:18.
37. Stone NJ, Robinson JG, Lichtenstein AH, Merz CNB, Blum CB, Eckel RH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association task force on practice guidelines. J Am Coll Cardiol 2014;63(25 Pt B):2889-2934 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.002] [Medline: 24239923]
38. Krumholz HM. Treatment of cholesterol in 2017. JAMA 2017;318(5):417-418 [FREE Full text] [doi: 10.1001/jama.2017.6753] [Medline: 28738130]
39. Charnock D, Shepperd S. Discern Online. United Kingdom: Radcliffe Online; 2021. URL: http://www.discern.org.uk/background_to_discern.php [accessed 2021-09-03]
40. Beaunoyer E, Arsenault M, Lomanowska AM, Guitton MJ. Understanding online health information: evaluation, tools, and strategies. Patient Educ Couns 2017;100(2):183-189 [FREE Full text] [doi: 10.1016/j.pec.2016.08.028] [Medline: 27595436]
41. Peabody D. Two components in bipolar scales: direction and extremeness. Psychol Rev 1962;69:65-73. [doi: 10.1037/h0039737] [Medline: 14484721]

42. Komorita SS, Graham WK. Number of scale points and the reliability of scales. Educ Psychol Meas 2016;25(4):987-995 [FREE Full text] [doi: 10.1177/001316446502500404]
43. Dolnicar S, Grün B, Leisch F, Rossiter J. Three good reasons NOT to use five and seven point Likert items. 2011 Presented at: 21st CAUTHE National Conference; February 8-11, 2011; Adelaide, Australia URL: https://ro.uow.edu.au/commpapers/775/
44. Weymann N, Härter M, Dirmaier J. Quality of online information on type 2 diabetes: a cross-sectional study. Health Promot Int 2015;30(4):821-831 [FREE Full text] [doi: 10.1093/heapro/dau019] [Medline: 24688114]
45. Freeman KS, Spyridakis JH. An examination of factors that affect the credibility of online health information. Tech Commun 2004;51(2):239-263 [FREE Full text]
46. Sillence E, Briggs P, Harris P, Fishwick L. A framework for understanding trust factors in web-based health advice. Int J Hum Comput Stud 2006;64(8):697-713 [FREE Full text] [doi: 10.1016/j.ijhcs.2006.02.007]
47. Neumark Y, Flum L, Lopez-Quintero C, Shtarkshall R. Quality of online health information about oral contraceptives from Hebrew-language websites. Isr J Health Policy Res 2012;1(1):38 [FREE Full text] [doi: 10.1186/2045-4015-1-38] [Medline: 23006798]
48. Metzger MJ. Making sense of credibility on the web: models for evaluating online information and recommendations for future research. J Am Soc Inf Sci Technol 2007;58(13):2078-2091. [doi: 10.1002/asi.20672]
49. Fitzsimmons PR, Michael BD, Hulley JL, Scott GO. A readability assessment of online Parkinson's disease information. J R Coll Physicians Edinb 2010;40(4):292-296. [doi: 10.4997/JRCPE.2010.401] [Medline: 21132132]
50. McLaughlin GH. SMOG grading-a new readability formula. J Read 1969;12(8):639-646 [FREE Full text]
51. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. Res Social Adm Pharm 2013;9(5):503-516 [FREE Full text] [doi: 10.1016/j.sapharm.2012.05.009] [Medline: 22835706]
52. Oliffe M, Thompson E, Johnston J, Freeman D, Bagga H, Wong PKK. Assessing the readability and patient comprehension of rheumatology medicine information sheets: a cross-sectional health literacy study. BMJ Open 2019;9(2):e024582 [FREE Full text] [doi: 10.1136/bmjopen-2018-024582] [Medline: 30813117]
53. Cheng C, Dunn M. Health literacy and the internet: a study on the readability of Australian online health information. Aust N Z J Public Health 2015;39(4):309-314 [FREE Full text] [doi: 10.1111/1753-6405.12341] [Medline: 25716142]
54. My Byline media. Readability Formulas. URL: https://readabilityformulas.com/ [accessed 2021-09-04]
55. MedicineNet. URL: https://www.medicinenet.com/statins/drug-class.htm#what_are_statins_and_how_do_they_work [accessed 2024-02-07]
56. Statins. Healthdirect. URL: https://www.healthdirect.gov.au/statins [accessed 2024-02-07]
57. Statin. Wikipedia. URL: https://en.wikipedia.org/wiki/Statin [accessed 2024-02-07]
58. Bruce DF. Statin side effects. WebMD. URL: https://www.webmd.com/cholesterol-management/side-effects-of-statin-drugs#1 [accessed 2024-02-07]
59. Mann D, Pallarito K. Are statins bad for you? 4 things to know. Health. URL: https://www.health.com/health/gallery/0,,20552165,00.html [accessed 2024-02-07]
60. Statins. MedlinePlus. URL: https://medlineplus.gov/statins.html [accessed 2024-02-07]
61. Statins. HealthEngine. URL: https://healthengine.com.au/info/statins [accessed 2024-02-07]
62. Statins. NHS. URL: https://www.nhs.uk/conditions/statins/ [accessed 2024-02-07]
63. Statin side effects: weigh the benefits and risks. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/in-depth/statin-side-effects/art-20046013 [accessed 2024-02-07]
64. Side effects of atorvastatin. Healthline. URL: https://www.healthline.com/health/high-cholesterol/statins-list-of-common-types [accessed 2024-02-07]
65. The uses and risks of statins. Medical News Today. URL: https://www.medicalnewstoday.com/articles/8274.php [accessed 2024-02-07]
66. Mercola J. You've been genetically hijacked to produce Pfizer's Pfrankenstein proteins. Mercola. 2024. URL: https://articles.mercola.com/sites/articles/archive/2010/07/20/the-truth-about-statin-drugs-revealed.aspx [accessed 2024-02-07]
67. Atorvastatin. NHS. URL: https://www.nhs.uk/medicines/atorvastatin/ [accessed 2024-02-07]
68. Statin adverse effects. Statin Effects Study. URL: https://statineffects.com/info/adverse_effects.htm [accessed 2024-02-07]
69. Getting treatment: the most widely used medicine to lower cholesterol is a statin, but there are other medicines available too and some may only be prescribed in a specialist lipid clinic. Heart UK. URL: https://www.heartuk.org.uk/getting-treatment/treatment-intro [accessed 2024-02-07]
70. Gajraj H. Why I've ditched statins for good. The Telegraph. 2014. URL: https://www.telegraph.co.uk/news/health/10717431/Why-Ive-ditched-statins-for-good.html [accessed 2024-02-07]
71. Statin medicines: frequently asked questions. NPS MedicineWise. URL: https://www.nps.org.au/consumers/statin-medicines-frequently-asked-questions [accessed 2024-02-07]
72. Zocor. RxList. URL: https://www.rxlist.com/zocor-side-effects-drug-center.htm [accessed 2024-02-07]
73. Simvastatin. Drugs.com. 2023. URL: https://www.drugs.com/simvastatin.html [accessed 2024-02-07]

74. Nedelman M. Statins: what happens to patients who stop taking them after side effects? CNN Health. 2017. URL: https://edition.cnn.com/2017/07/24/health/statins-side-effects-outcomes-study/index.html [accessed 2024-02-07]
75. Bernard S, Cooke T, Cole T, Hachani L, Bernard J. Quality and readability of online information about type 2 diabetes and nutrition. JAAPA 2018;31(11):41-44 [FREE Full text] [doi: 10.1097/01.JAA.0000546481.02560.4e] [Medline: 30358679]
76. How search algorithms work mountain view. Google. URL: https://www.google.com/search/howsearchworks/algorithms/ [accessed 2023-12-29]
77. Makino T, Jung C, Phan D. Finding more mobile-friendly search results. Google Search Central Blog. 2015. URL: https://developers.google.com/search/blog/2015/02/finding-more-mobile-friendly-search] [accessed 2023-12-29]
78. Webmaster guidelines mountain view. Google. 2021. URL: https://developers.google.com/search/docs/essentials [accessed 2023-12-29]
79. Weiss BD. Progress in health literacy-encouraging, and not so encouraging. Health Lit Res Pract 2017;1(4):e231-e232 [FREE Full text] [doi: 10.3928/24748307-20171016-01] [Medline: 31294267]
80. Weiss BD, Schwartzberg JG. Health Literacy and Patient Safety: Help Patients Understand: Manual for Clinicians, Volume 1. Tucson: American Medical Association; 2007.
81. Kasabwala K, Agarwal N, Hansberry DR, Baredes S, Eloy JA. Readability assessment of patient education materials from the American Academy of Otolaryngology—Head and Neck Surgery Foundation. Otolaryngol Head Neck Surg 2012;147(3):466-471 [FREE Full text] [doi: 10.1177/0194599812442783] [Medline: 22473833]
82. Breakey VR, Harris L, Davis O, Agarwal A, Ouellette C, Akinnawo E, et al. The quality of information about sickle cell disease on the internet for youth. Pediatr Blood Cancer 2017;64(4):e26309 [FREE Full text] [doi: 10.1002/pbc.26309] [Medline: 27786409]


## Abbreviations

**CVD:** cardiovascular disease
**HONcode:** Health on the Net Code of Conduct
**SMOG:** Simple Measure of Gobbledygook


*Edited by G Eysenbach; submitted 21.09.22; peer-reviewed by L Daraz, WF Khaw; comments to author 24.12.22; revised version received 13.04.23; accepted 17.11.23; published 14.03.24.*

*Please cite as:*
*Ling E, de Pieri D, Loh E, Scott KM, Li SCH, Medbury HJ*
*Evaluation of the Accuracy, Credibility, and Readability of Statin-Related Websites: Cross-Sectional Study*
*Interact J Med Res 2024;13:e42849*
*URL: https://www.i-jmr.org/2024/1/e42849*
*doi:10.2196/42849*
*PMID:38483461*

Original Paper

# Online Visibility and Scientific Relevance of Strabismus Research: Bibliometric Analysis

Aleksander Stupnicki[1]; Basil Suresh[1], iBSc (Hons), MBBS; Saurabh Jain[1,2], MB BS

[1]University College London Medical School, London, United Kingdom

[2]Royal 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.

***(Interact J Med Res 2024;13:e50698)*** doi:10.2196/50698

**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 AASs (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

normalized citations or IF were not significant at $P<.05$ (Multimedia Appendix 1).

### Multivariate 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 - ijmr_v13i1e50698_app1.xlsx ]

## References

1. Surani Z, Hirani R, Elias A, Quisenberry L, Varon J, Surani S, et al. Social media usage among health care providers. BMC Res Notes 2017;10(1):654 [FREE Full text] [doi: 10.1186/s13104-017-2993-y] [Medline: 29187244]
2. Hashemi H, Pakzad R, Heydarian S, Yekta A, Aghamirsalim M, Shokrollahzadeh F, et al. Global and regional prevalence of strabismus: a comprehensive systematic review and meta-analysis. Strabismus 2019;27(2):54-65. [doi: 10.1080/09273972.2019.1604773] [Medline: 31012389]
3. Thompson DF, Walker CK. A descriptive and historical review of bibliometrics with applications to medical sciences. Pharmacotherapy 2015;35(6):551-559. [doi: 10.1002/phar.1586] [Medline: 25940769]
4. García-Villar C. A critical review on altmetrics: can we measure the social impact factor? Insights Imaging 2021;12(1):92 [FREE Full text] [doi: 10.1186/s13244-021-01033-2] [Medline: 34215935]
5. Priem J, Hemminger BH. Scientometrics 2.0: new metrics of scholarly impact on the social Web. First Monday 2010;15(7) [FREE Full text] [doi: 10.5210/fm.v15i7.2874]
6. Boudry C, Baudouin C, Mouriaux F. International publication trends in dry eye disease research: a bibliometric analysis. Ocul Surf 2018;16(1):173-179. [doi: 10.1016/j.jtos.2017.10.002] [Medline: 29031646]
7. Patel PA, Patel PN, Becerra AZ, Mehta MC. Bibliometric analysis of the 100 most-disruptive articles in ophthalmology. Clin Exp Ophthalmol 2022;50(6):690-695. [doi: 10.1111/ceo.14109] [Medline: 35593696]
8. Huang A, Kim S, Zhu H, Pathare N, Ooi XY, Kirby RP, et al. A bibliometric analysis of ophthalmology resident research productivity in the United States. medRxiv Preprint posted online on February 08, 2022. [doi: 10.1101/2022.02.06.22270335]
9. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc 2018;106(4):531-541 [FREE Full text] [doi: 10.5195/jmla.2018.283] [Medline: 30271302]
10. Strabismus. Cambridge Dictionary. 2024. URL: https://dictionary.cambridge.org/dictionary/english/strabismus [accessed 2024-03-27]

11. Advanced search. Altmetric. URL: https://help.altmetric.com/support/solutions/articles/6000240857-advanced-search [accessed 2024-05-09]
12. Myers BA, Kahn KL. Practical publication metrics for academics. Clin Transl Sci 2021;14(5):1705-1712 [FREE Full text] [doi: 10.1111/cts.13067] [Medline: 33982433]
13. Ellegaard O, Wallin JA. The bibliometric analysis of scholarly production: how great is the impact? Scientometrics 2015;105(3):1809-1831 [FREE Full text] [doi: 10.1007/s11192-015-1645-z]
14. Thelwall M, Wilson P. Regression for citation data: an evaluation of different methods. J Informetr 2014;8(4):963-971 [FREE Full text] [doi: 10.1016/j.joi.2014.09.011]
15. Taherdoost H, Sahibuddin S, Jalaliyoon N. Exploratory factor analysis; concepts and theory. SSRN J 2022 [FREE Full text]
16. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
17. Thelwall M, Nevill T. Could scientists use Altmetric.com scores to predict longer term citation counts? J Informetr 2018;12(1):237-248 [FREE Full text] [doi: 10.1016/j.joi.2018.01.008]
18. Haustein S, Larivière V, Thelwall M, Amyot D, Peters I. Tweets vs. Mendeley readers: how do these two social media metrics differ? Inf Technol 2014;56(5):207-215 [FREE Full text] [doi: 10.1515/itit-2014-1048]
19. Yu H, Murat B, Li J, Li L. How can policy document mentions to scholarly papers be interpreted? An analysis of the underlying mentioning process. Scientometrics 2023;128:6247-6266 [FREE Full text] [doi: 10.1007/s11192-023-04826-y]
20. Elmore SA. The Altmetric attention score: what does it mean and why should i care? Toxicol Pathol 2018;46(3):252-255 [FREE Full text] [doi: 10.1177/0192623318758294] [Medline: 29448902]


## Abbreviations

**AAS:** Altmetric Attention Score
**IF:** impact factor
**WoS:** Web of Science


*Edited by T de Azevedo Cardoso; submitted 10.07.23; peer-reviewed by A Santisteban, Y Jiang; comments to author 01.02.24; revised version received 27.03.24; accepted 29.04.24; published 12.06.24.*

*Please cite as:*
*Stupnicki A, Suresh B, Jain S*
*Online Visibility and Scientific Relevance of Strabismus Research: Bibliometric Analysis*
*Interact J Med Res 2024;13:e50698*
*URL: https://www.i-jmr.org/2024/1/e50698*
*doi:10.2196/50698*
*PMID:38865170*

Original Paper

# Census-Dependent Mortality of Ventilated Patients With COVID-19 in Israel: Noninterventional Observational Cohort Study

Joseph Mendlovic[1,2], MHA, MD; Francis B Mimouni[3,4], MD

[1]Shaare Zedek Medical Center, Affiliated with the Hadassah-Hebrew University School of Medicine, Jerusalem, Israel

[2]Associate Director General Office, Israel Ministry of Health, Jerusalem, Israel

[3]Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[4]Leumit 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.

***(Interact J Med Res 2024;13:e41749)*** doi:10.2196/41749

**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 $\alpha$ 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).

| | Age[a] (years), mean (SD; median; IQR) | Gender[b] (male:female), n:n (% males) | Length of stay to outcome (days)[a], mean (SD; median; IQR) | Length of stay to ventilation (days)[a], mean (SD; median; IQR) | Length of stay from ventilation to outcome (days)[a], mean (SD; median; IQR) |
|---|---|---|---|---|---|
| Alive (n=1128; 33.44%) | 60.3 (16.3; 62; 51-71) | 721:407 (63.9%) | 35.7 (29.4; 27; 15-47) | 4.9 (5.9; 3; 1.6) | 30.8 (28.4; 22; 11-42) |
| Dead (n=2245; 66.56%) | 72.2 (13.0; 74; 65-82) | 1487:758 (66.2%) | 18.7 (16.5; 15; 8-25) | 6.0 (6.6; 4; 1-8) | 12.8 (15.1; 8; 3-17) |

[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.

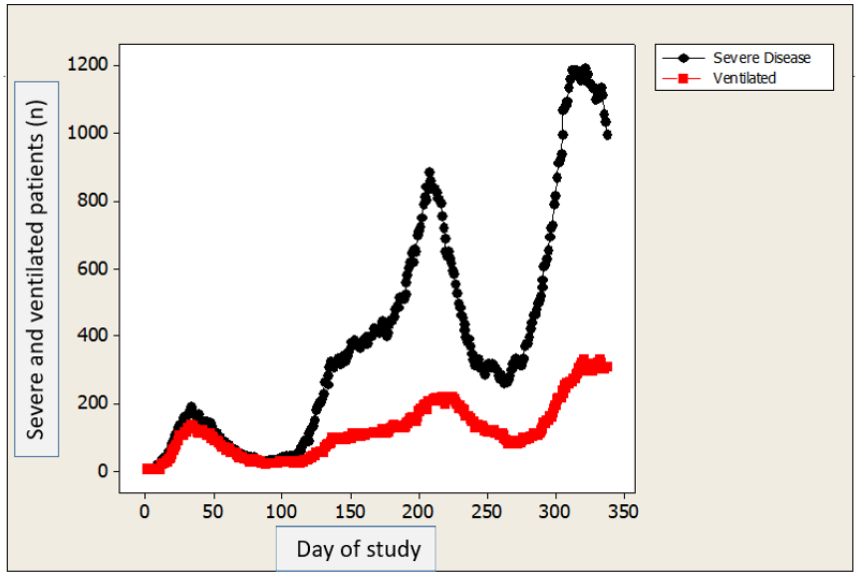


**Figure 2.** Daily mortality over time.

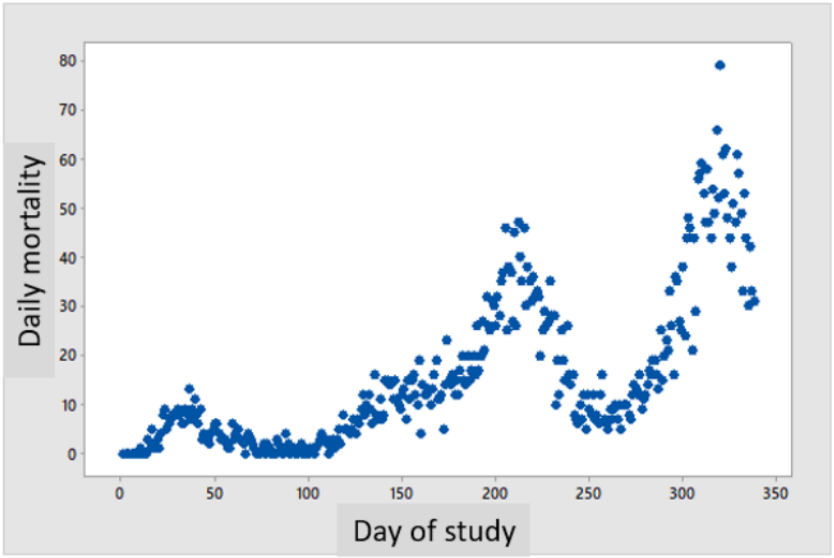



**Figure 3.** Daily percent mortality of ventilated patients over time.

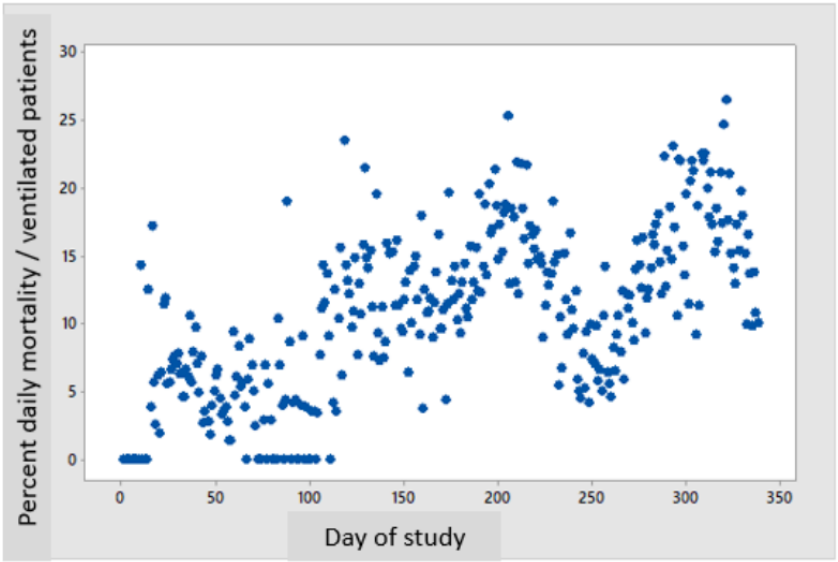



**Figure 4.** Daily mortality rates versus the daily census of ventilated patients.

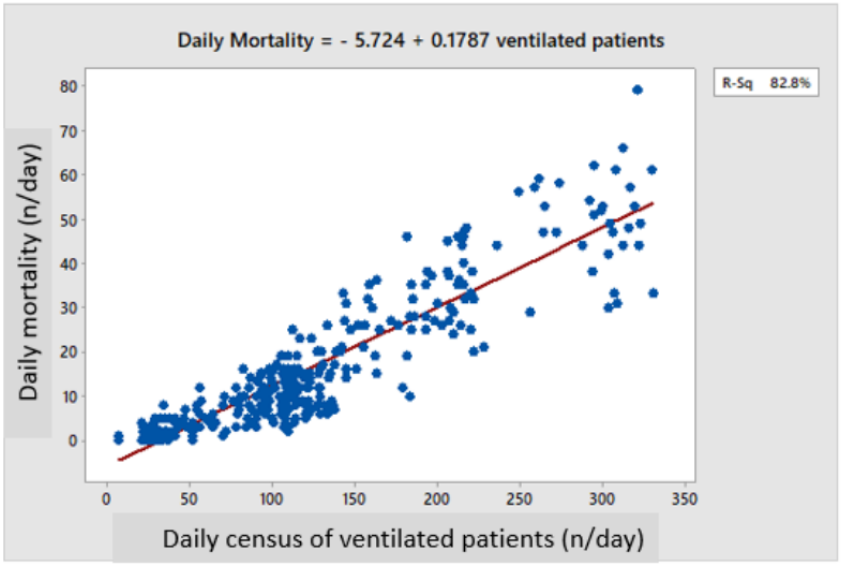


**Figure 5.** Daily percent mortality of ventilated patients versus daily census of ventilated patients.

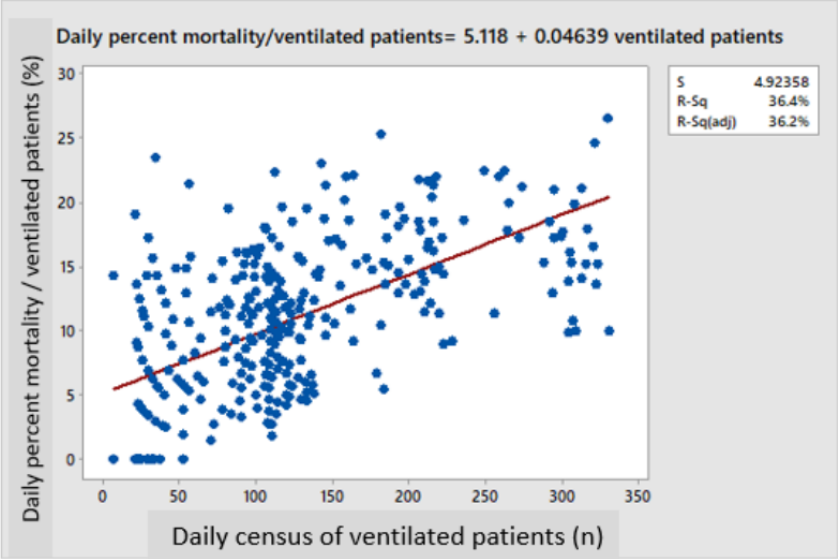


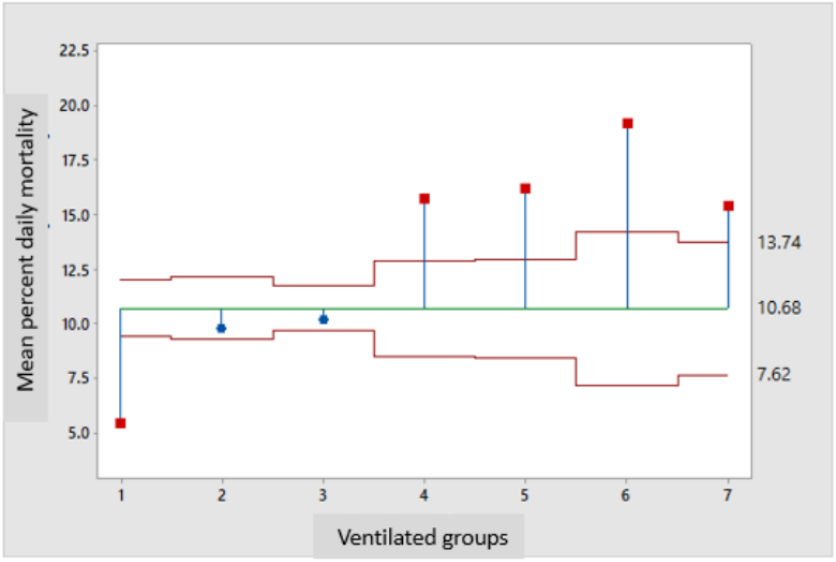

**Table 2.** Multiple regression analysis results showing the relative contribution of the daily census of ventilated patients, gender, and age (independent variables) on the daily percent mortality of ventilated patients (dependent variable).

| | Daily census of ventilated patients[a] | Gender[a] | Age[a] | Model summary |
|---|---|---|---|---|
| Partial $R^2$ | 0.3574 | 0.0275 | 0.0005 | 0.3915 |
| Effect size (adjusted sum of squares) | 3772.7 | 290.7 | 68.6 | 5012.5 |
| *P* value | <.001 | <.001 | .09 | <.001 |

[a]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

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 [accessed 2021-02-25]
2. About variants of the virus that causes COVID-19. World Health Organization. 2021 Feb 21. URL: https://www.cdc.gov/coronavirus/2019-ncov/transmission/variant.html [accessed 2021-02-21]
3. Maves RC, Downar J, Dichter JR, Hick JL, Devereaux A, Geiling JA, ACCP Task Force for Mass Critical Care. Triage of scarce critical care resources in COVID-19 an implementation guide for regional allocation: an expert panel report of the task force for mass critical care and the American College of Chest Physicians. Chest 2020 Jul;158(1):212-225 [FREE Full text] [doi: 10.1016/j.chest.2020.03.063] [Medline: 32289312]
4. Rosenbaum L. Facing Covid-19 in Italy - ethics, logistics, and therapeutics on the epidemic's front line. N Engl J Med 2020 May 14;382(20):1873-1875. [doi: 10.1056/NEJMp2005492] [Medline: 32187459]
5. Critical preparedness, readiness and response actions for COVID-19. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/critical-preparedness-readiness-and-response-actions-for-covid-19 [accessed 2021-02-25]
6. Wang X, Zhang X, He J. Challenges to the system of reserve medical supplies for public health emergencies: reflections on the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic in China. Biosci Trends 2020 Mar 16;14(1):3-8 [FREE Full text] [doi: 10.5582/bst.2020.01043] [Medline: 32062645]
7. Haklai Z, Aburbeh M, Goldberger N, Gordon E. Excess mortality during the COVID-19 pandemic in Israel, March-November 2020: when, where, and for whom? Isr J Health Policy Res 2021 Feb 26;10(1):17 [FREE Full text] [doi: 10.1186/s13584-021-00450-4] [Medline: 33637126]
8. Horton R. Offline: COVID-19 and the NHS-"a national scandal". Lancet 2020 Mar 28;395(10229):1022 [FREE Full text] [doi: 10.1016/S0140-6736(20)30727-3] [Medline: 32222186]
9. Janke AT, Mei H, Rothenberg C, Becher RD, Lin Z, Venkatesh AK. Analysis of hospital resource availability and COVID-19 mortality across the United States. J Hosp Med 2021 Apr;16(4):211-214 [FREE Full text] [doi: 10.12788/jhm.3539] [Medline: 33496664]
10. Sen-Crowe B, Sutherland M, McKenney M, Elkbuli A. A closer look into global hospital beds capacity and resource shortages during the COVID-19 pandemic. J Surg Res 2021 Apr;260:56-63 [FREE Full text] [doi: 10.1016/j.jss.2020.11.062] [Medline: 33321393]
11. The lack of intensive care beds is general hospitals in Israel [Article in Hebrew]. State of Israel, Parliament's Research and Information Department. URL: https://fs.knesset.gov.il/globaldocs/MMM/aef26d8d-f1f7-e411-80c8-00155d01107c/2_aef26d8d-f1f7-e411-80c8-00155d01107c_11_9172.pdf [accessed 2021-03-26]
12. Adhikari NK, Fowler RA, Bhagwanjee S, Rubenfeld GD. Critical care and the global burden of critical illness in adults. Lancet 2010 Oct;376(9749):1339-1346. [doi: 10.1016/s0140-6736(10)60446-1]
13. Coronavirus. State of Israel, Ministry of Health. 2021. URL: https://www.gov.il/en/departments/topics/corona-main-sub [accessed 2021-03-08]
14. Admissions. State of Israel, Ministry of Health. 2021. URL: https://www.gov.il/he/pages/general-hospitalization [accessed 2021-03-08]
15. Simchen E, Sprung CL, Galai N, Zitser-Gurevich Y, Bar-Lavi Y, Levi L, et al. Survival of critically ill patients hospitalized in and out of intensive care. Crit Care Med 2007 Feb;35(2):449-457. [doi: 10.1097/01.CCM.0000253407.89594.15] [Medline: 17167350]
16. Final Coronavirus Disease (COVID-19) Treatment Guidelines (February 29, 2024). National Institutes of Health. URL: https://www.covid19treatmentguidelines.nih.gov/ [accessed 2020-03-09]
17. Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. N Engl J Med 2020 Dec 17;383(25):2451-2460. [doi: 10.1056/NEJMcp2009575] [Medline: 32412710]
18. 45 CFR 46. US Department of Health and Human Services. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html [accessed 2024-06-11]
19. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. Lancet Glob Health 2020 Apr;8(4):e480 [FREE Full text] [doi: 10.1016/S2214-109X(20)30068-1] [Medline: 32109372]
20. Dale CR, Starcher RW, Chang SC, Robicsek A, Parsons G, Goldman JD, et al. Surge effects and survival to hospital discharge in critical care patients with COVID-19 during the early pandemic: a cohort study. Crit Care 2021 Feb 17;25(1):70 [FREE Full text] [doi: 10.1186/s13054-021-03504-w] [Medline: 33596975]
21. Staff T. COVID patient dies after breathing tube detaches, staff fails to notice in time. The Times of Israel. 2021. URL: https://www.timesofisrael.com/covid-patient-dies-after-breathing-tube-detaches-staff-fails-to-notice-in-time/ [accessed 2021-02-25]
22. Bridgwater A. How Israel became a technology startup nation. Forbes. 2020. URL: https://www.forbes.com/sites/adrianbridgwater/2020/02/21/how-israel-became-a-technology-startup-nation/?sh=2039274b780e [accessed 2021-03-18]
23. Building a resilient recovery: emerging stronger from the COVID-19 pandemic. OECD. URL: https://www.oecd.org/coronavirus/en [accessed 2021-02-25]

24. Staff T. Hospitals reportedly told to prep to reuse protective gear amid shortage fears. The Times of Israel. 2020. URL: https://www.timesofisrael.com/hospitals-reportedly-told-to-prep-to-reuse-protective-gear-amid-shortage-fears/ [accessed 2024-06-11]
25. Lim ZJ, Subramaniam A, Ponnapa Reddy M, Blecher G, Kadam U, Afroz A, et al. Case fatality rates for patients with COVID-19 requiring invasive mechanical ventilation. A meta-analysis. Am J Respir Crit Care Med 2021 Jan 01;203(1):54-66 [FREE Full text] [doi: 10.1164/rccm.202006-2405OC] [Medline: 33119402]
26. Heyl J, Hardy F, Tucker K, Hopper A, Marchã MJM, Navaratnam AV, et al. Frailty, comorbidity, and associations with in-hospital mortality in older COVID-19 patients: exploratory study of administrative data. Interact J Med Res 2022 Dec 12;11(2):e41520 [FREE Full text] [doi: 10.2196/41520] [Medline: 36423306]
27. Uchino K, Kolikonda MK, Brown D, Kovi S, Collins D, Khawaja Z, et al. Decline in stroke presentations during COVID-19 surge. Stroke 2020 Aug;51(8):2544-2547 [FREE Full text] [doi: 10.1161/STROKEAHA.120.030331] [Medline: 32716818]
28. De Filippo O, D'Ascenzo F, Angelini F, Bocchino PP, Conrotto F, Saglietto A, et al. Reduced rate of hospital admissions for ACS during Covid-19 outbreak in Northern Italy. N Engl J Med 2020 Jul 02;383(1):88-89 [FREE Full text] [doi: 10.1056/NEJMc2009166] [Medline: 32343497]
29. Maleitzke T, Pumberger M, Gerlach UA, Herrmann C, Slagman A, Henriksen LS, et al. Impact of the COVID-19 shutdown on orthopedic trauma numbers and patterns in an academic Level I Trauma Center in Berlin, Germany. PLoS One 2021;16(2):e0246956 [FREE Full text] [doi: 10.1371/journal.pone.0246956] [Medline: 33592046]
30. COVIDSurg Collaborative. Elective surgery cancellations due to the COVID-19 pandemic: global predictive modelling to inform surgical recovery plans. Br J Surg 2020 Oct;107(11):1440-1449 [FREE Full text] [doi: 10.1002/bjs.11746] [Medline: 32395848]
31. Patel BV, Haar S, Handslip R, Auepanwiriyakul C, Lee TM, Patel S, United Kingdom COVID-ICU National Service Evaluation. Natural history, trajectory, and management of mechanically ventilated COVID-19 patients in the United Kingdom. Intensive Care Med 2021 May;47(5):549-565 [FREE Full text] [doi: 10.1007/s00134-021-06389-z] [Medline: 33974106]
32. Ministry of Health's World of Data. URL: https://datadashboard.health.gov.il/portal/dashboard/corona [accessed 2024-06-13]


## Abbreviations

**ICU:** intensive care unit
**LOS:** length of stay
**MOH:** Ministry of Health
**OECD:** Organisation for Economic Co-operation and Development


*Edited by T de Azevedo Cardoso; submitted 07.08.22; peer-reviewed by S De Leo, B Sainz-de-Abajo, S Pesälä, C Zhao; comments to author 14.11.23; revised version received 06.12.23; accepted 20.05.24; published 09.07.24.*

*Please cite as:*
*Mendlovic J, Mimouni FB*
*Census-Dependent Mortality of Ventilated Patients With COVID-19 in Israel: Noninterventional Observational Cohort Study*
*Interact J Med Res 2024;13:e41749*
*URL: https://www.i-jmr.org/2024/1/e41749*
*doi:10.2196/41749*
*PMID:*

Original Paper

# Chinese Health Insurance in the Digital Era: Bibliometric Study

Zhiyuan Hu[1], BSc; Xiaoping Qin[1], MPH; Kaiyan Chen[2], BMedSci; Yu-Ni Huang[3], PhD; Richard Szewei Wang[4], MSc; Tao-Hsin Tung[5], PhD; Yen-Ching Chuang[6], PhD; Bing-Long Wang[1], PhD, MBA

[1]School of Health Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]Department of Education, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[3]College of Medical and Health Science, Asia University, Taichung, Taiwan

[4]Affiliation Program of Data Analytics and Business Computing, Stern School of Business, New York University, New York, NY, United States

[5]Evidence-based Medicine Center, Taizhou Hospital of Zhejiang Province affiliated to Wenzhou Medical University, Taizhou, China

[6]Business 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 $R^2$=0.6788. In total, 62 institutions and 811 authors have published research articles in the Chinese language in this field. 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.

***(Interact J Med Res 2024;13:e52020)*** doi:10.2196/52020

**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 health 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

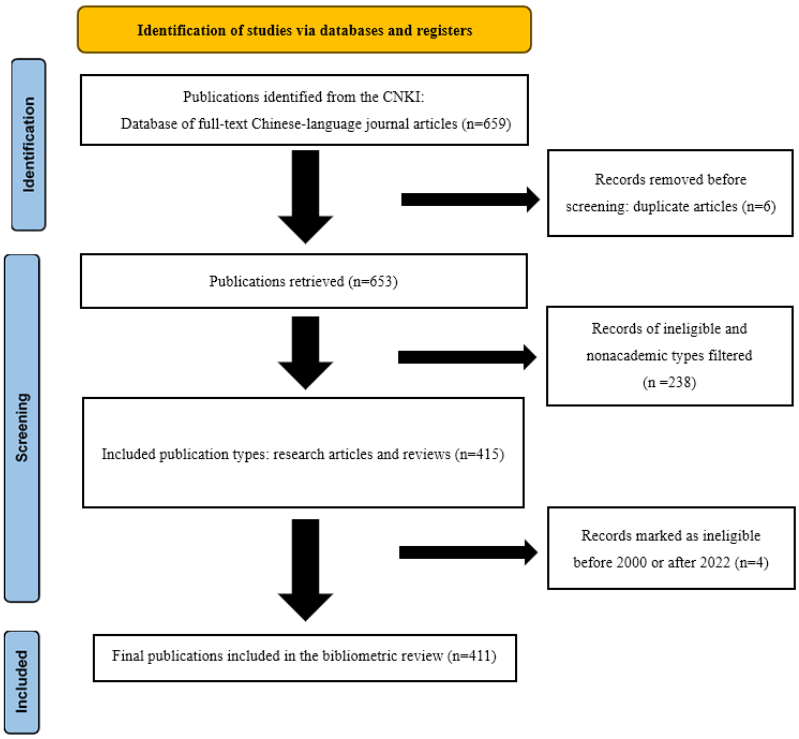
why the beginning year was set as 2000 is that China did not formally establish its current concept of a health insurance system until 1999 [1,2].

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.

| Rank | Journal | Articles, n (%) |
|---|---|---|
| 1 | *China Health Insurance* | 28 (6.8) |
| 2 | *China Digital Medicine* | 20 (4.8) |
| 3 | *Health Economics Research* | 17 (4.1) |
| 4 | *Chinese Hospitals* | 15 (3.6) |
| 5 | *China Social Security* | 14 (3.4) |
| 6 | *Journal of Medical Informatics* | 11 (2.6) |
| 7 | *Chinese Journal of Health Informatics and Management* | 11 (2.6) |
| 8 | *China Health* | 9 (2.1) |
| 9 | *Modern Hospital* | 8 (1.9) |
| 10 | *Chinese Health Service Management* | 8 (1.9) |

## 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.

| Rank | Title | Journal | Authors | Citations, n |
|---|---|---|---|---|
| 1 | Internet + Medical Mode： Contents and System Architecture | *Chinese Hospital Management* | Zhu Jinsong | 201 |
| 2 | Research on Development Policy of Integration of Medical Care and Pension and Institutional Pension Service for the Elderly | *Medicine and Society* | Ma Lili, Chen Na, and Tang Shaoliang | 194 |
| 3 | The Status Quo of Internet Medical-Based on The Analysis And Investigation on Three Hospitals | *Chinese Journal of Health Policy* | Wang Anqi and Zheng Xueqian | 138 |
| 4 | Analysis on the Development Model of Internet Hospitals in China | *Health Economics Research* | Zhang Mengqian, Wang Yancui, Qian Zhenguang, and Wang Dandan | 59 |
| 5 | Practice and Exploration of Medical Association in Remote Areas | *Modern Hospital Management* | Sun Xizhuo, Gong Fangfang, Gu Xiaodong, Su Qian, and Cai Yutong | 55 |
| 6 | Problems and Countermeasures for “Internet + Healthcare” in China | *Administration Reform* | Luan Yunbo and Tian Zhendu | 52 |
| 7 | Problems and Countermeasures for Medical Service Supply in Elderly Care Institutions from the Perspective of Medical-Old-Age Combination | *Chinese Journal of Gerontology* | Fan Qingmei, Chen Le, Wu Meng, Wu Jiankang, and Li Jiamin | 45 |
| 8 | Analysis on Problems and Countermeasures of Mobile Health Service in China | *Medicine and Philosophy* | Yang Xiaoli and Feng Xinwei | 45 |
| 9 | Study on Regulation System and Related Mechanism of Internet Based Medicine | *Chinese Journal of Health Informatics and Management* | Meng Qun, Yin Xin, and Dong Kenan | 44 |
| 10 | Analysis on Service Mode and Application Status for Network Medical Treatment in China | *Chinese Journal of Health Informatics and Management* | Liu Ning and Chen Min | 42 |

## 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.

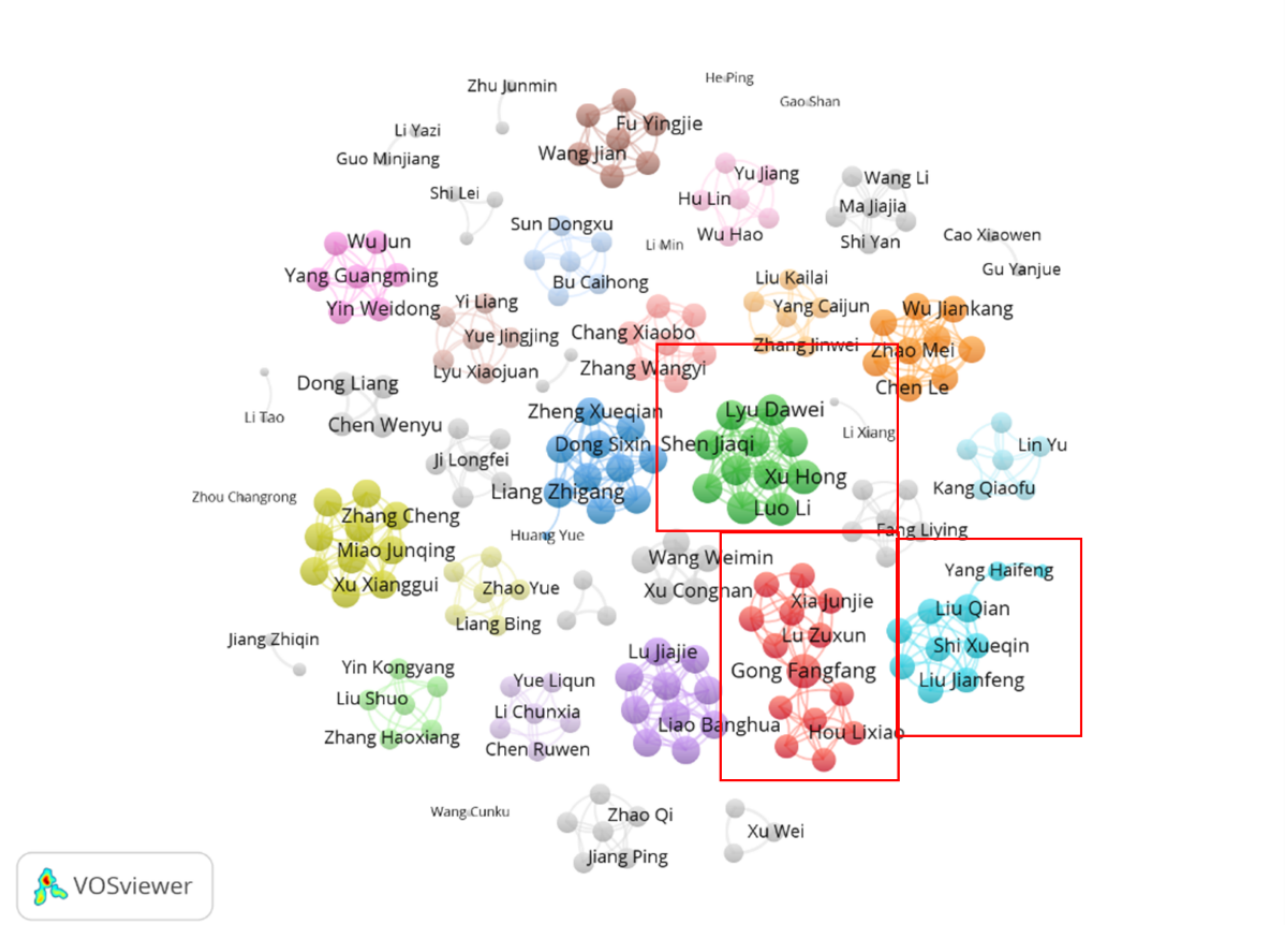



**Table 3.** Top Chinese authors ranked according to total link strength.

| Rank | Author | Total link strength |
|---|---|---|
| 1 | Xu Hong | 16 |
| 2 | Lyu Dawei | 16 |
| 3 | Zheng Xueqian | 14 |
| 4 | Luo Li | 14 |
| 5 | Wang Weijun | 14 |
| 6 | Liu Qian | 14 |
| 7 | Gong Fangfang | 13 |
| 8 | Sun Xizhuo | 13 |
| 9 | Liang Zhigang | 12 |

## 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.

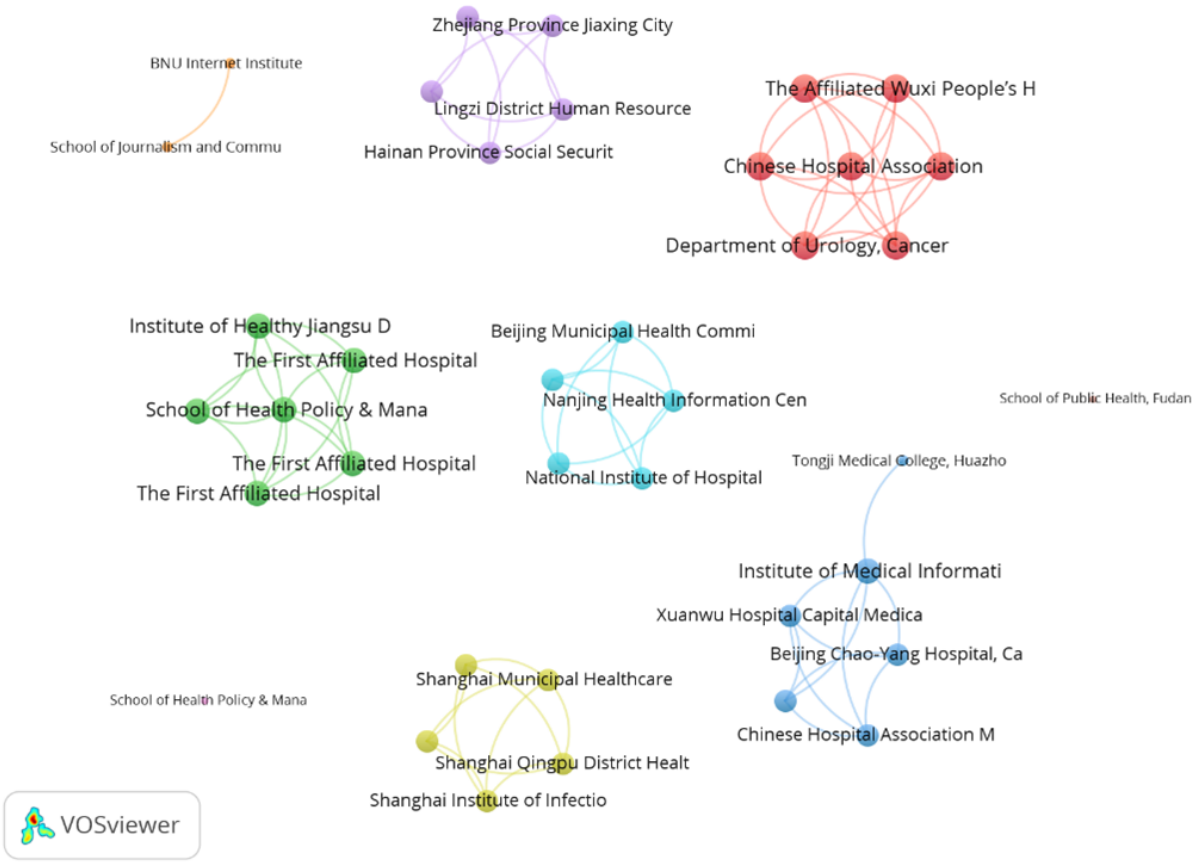



**Table 4.** Total link strength and institution co-occurrence of publications.

| Rank | Institution | Total link strength |
|---|---|---|
| 1 | School of Health Policy & Management, Nanjing Medical University | 6 |
| 2 | Chinese Hospital Association | 6 |
| 3 | Peking University Health Science Center | 6 |
| 4 | Department of Urology, Cancer Hospital Chinese Academy of Medical Sciences | 6 |
| 5 | Beijing Municipal Health Commission | 6 |
| 6 | Nantong University Medical School | 6 |
| 7 | The Second Hospital of Dalian Medical University, Department of Neurosurgery | 6 |
| 8 | Institute of Healthy Jiangsu Development, Nanjing Medical University | 6 |
| 9 | Shanghai Municipal Healthcare Security Bureau | 6 |
| 10 | Shanghai Institute of Infectious Disease and Biosecurity, School of Public Health of Fudan University | 6 |
| 11 | Chinese Hospital Association Medical Legality Specialized Committee | 6 |

## 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).

**Figure 4.** Visualization of keyword co-occurrence analysis. IPHC: Internet Plus Healthcare.

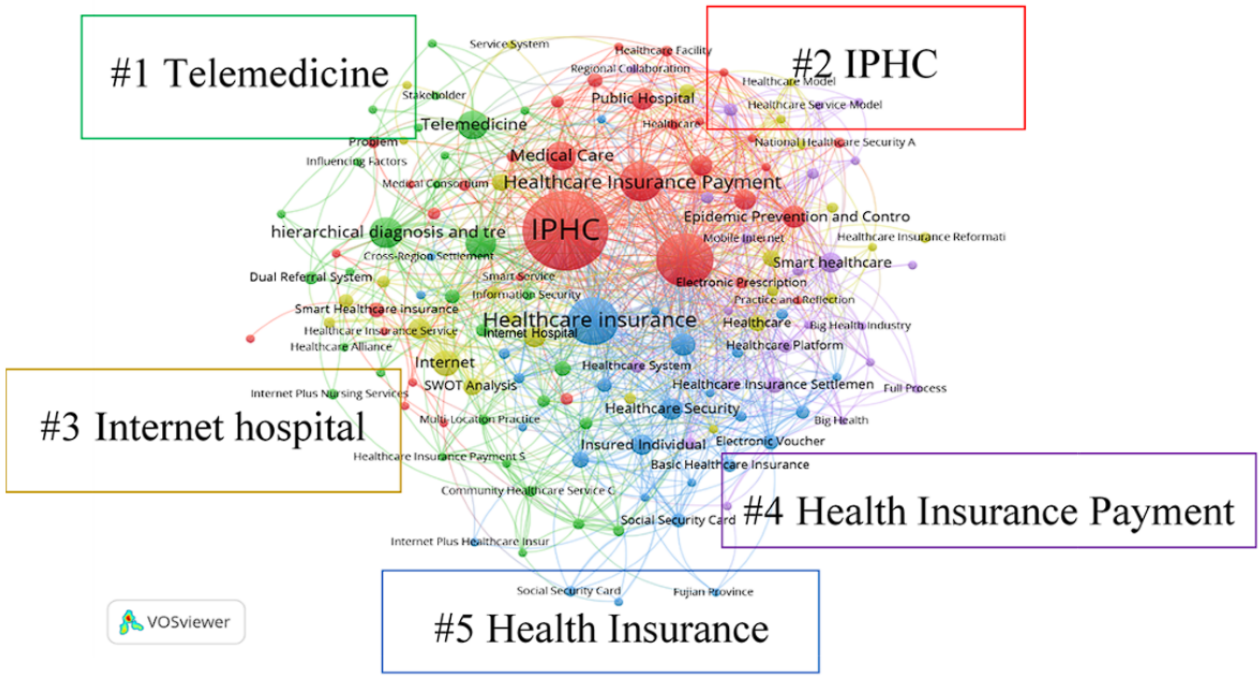



**Table 5.** Main keywords translated to English and their total link strength (TLS).

| Keywords | TLS |
|---|---|
| **Cluster 1: telemedicine** | |
| hierarchical diagnosis and treatment | 109 |
| telemedicine | 102 |
| healthcare alliance | 81 |
| **Cluster 2: IPHC[a]** | |
| internet healthcare | 129 |
| healthcare service | 68 |
| healthcare insurance fund | 79 |
| **Cluster 3: internet hospital** | |
| internet hospital | 179 |
| smart healthcare insurance | 29 |
| big data | 56 |
| **Cluster 4: insurance payment** | |
| online payment | 72 |
| healthcare | 31 |
| smart healthcare | 29 |
| **Cluster 5: health insurance** | |
| national health security administration | 41 |
| insured individual | 53 |
| social security card | 41 |

[a]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 interconnection 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 forefronts 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, XQ, 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.
[PNG File , 55 KB - ijmr_v13i1e52020_app1.png ]

## References


1. The Decision on Establishing the Basic Medical Insurance System for Urban Employees. The Central People's Government of the People's Republic of China. 1999. URL: http://www.gov.cn/banshi/2005-08/04/content_20256.htm [accessed 2024-07-06]
2. Jakovljevic M, Chang H, Pan J, Guo C, Hui J, Hu H, et al. Successes and challenges of China's health care reform: a four-decade perspective spanning 1985-2023. Cost Eff Resour Alloc 2023 Aug 30;21(1):59 [FREE Full text] [doi: 10.1186/s12962-023-00461-9] [Medline: 37649062]
3. Li Z, Li J. Lessons and prospects of Universal Health Coverage in China: the importance of equity, quality, and affordability. Asian Bioeth Rev 2019 Mar;11(1):21-40 [FREE Full text] [doi: 10.1007/s41649-019-00077-3] [Medline: 33717298]
4. Guiding Opinions of the National Healthcare Security Administration on Actively Promoting the Payment of Internet plus Healthcare by Health insurance [Article in Chinese]. The Central People's Government of the People's Republic of China. 2020. URL: http://www.gov.cn/zhengce/zhengceku/2020-11/03/content_5556883.htm [accessed 2024-06-07]
5. Ehrmeyer SS, Laessig RH. Total quality management in hematology. Accreditation and Quality Assurance 2000 Sep 5;5(9):360-362. [doi: 10.1007/s007690000188]
6. Yang F, Shu H, Zhang X. Understanding "Internet Plus Healthcare" in China: policy text analysis. J Med Internet Res 2021 Jul 26;23(7):e23779 [FREE Full text] [doi: 10.2196/23779] [Medline: 34309581]
7. Zhou X, Chen L. Digital health care in China and access for older people. Lancet Public Health 2021 Dec;6(12):e873-e874 [FREE Full text] [doi: 10.1016/S2468-2667(21)00051-7] [Medline: 34197810]
8. Huang L. Internet plus healthcare under the pressure of the epidemic. China Hospital CEO 2020 Mar 15;2020(05):37-41 (in Chinese) [FREE Full text]
9. Zhang S, Ju W, Shen T. The development prospect of internet medicine under the epidemic situation. China Digital Medicine 2020 Sep 15;15(09):15-17 (in Chinese) [FREE Full text] [doi: 10.3969/j.issn.1673-7571.2020.09.004]
10. Jeffer EK. Total quality management and the army health care system. Mil Med 1991 Oct;156(10):546-550. [Medline: 1749501]
11. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res 2020 Jul 29;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]
12. Ahmadvand A, Kavanagh D, Clark M, Drennan J, Nissen L. Trends and visibility of "Digital Health" as a keyword in articles by JMIR Publications in the new millennium: bibliographic-bibliometric analysis. J Med Internet Res 2019 Dec 19;21(12):e10477 [FREE Full text] [doi: 10.2196/10477] [Medline: 31855190]
13. Yang K, Qi H. Research on health disparities related to the COVID-19 pandemic: a bibliometric analysis. Int J Environ Res Public Health 2022 Jan 22;19(3) [FREE Full text] [doi: 10.3390/ijerph19031220] [Medline: 35162243]
14. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
15. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022 Jul 06;24(7):e35816 [FREE Full text] [doi: 10.2196/35816] [Medline: 35793141]
16. Ji L. The current status and research on strategies for the development of telemedicine. China Economist 2018 Dec 15(12):244-245 (in Chinese) [FREE Full text] [doi: 10.3969/j.issn.1004-4914.2018.12.123]
17. Lin X, Li J, Ti S, Qi J, Yang H. Problems and countermeasures in medical insurance payment for "Internet+" medical services. Chinese Journal of Medical Management Sciences 2021 Mar 23;11(02):35-37 (in Chinese) [FREE Full text] [doi: 10.3969/j.issn.2095-7432.2021.02.006]
18. Yuan J, Xiao Y, Shi Z, Liu X. Analysis and thinking of telemedicine to better meet the needs of patients with chronic disease. Chinese Hospitals 2021 Aug 14;25(08):44-47 (in Chinese) [FREE Full text] [doi: 10.19660/j.issn.1671-0592.2021.8.14]
19. Guo M, Liu Y. Research on development opportunities and countermeasures of “Internet Plus Chronic Disease Management”. China Health Insurance 2022;2022(07):46-52 [FREE Full text] [doi: 10.19546/j.issn.1674-3830.2022.7.008]

20. Ding T, Qian A, Tan Z. Comparative analysis of domestic and foreign internet medical service pricing and medical insurance payment policies. Chinese Hospitals 2022 Sep;26(09):10-13 (in Chinese) [FREE Full text] [doi: 10.19660/j.issn.1671-0592.2022.9.03]
21. Zheng D, Nie L, Wang L, Chen Y. Medical insurance management under the background of "Internet Plus Medical". Chinese Hospitals 2021;25(06):5-8 (in Chinese) [FREE Full text]
22. Zhou X, Zhou M, Huang D, Cui L. A probabilistic model for co-occurrence analysis in bibliometrics. J Biomed Inform 2022 Apr;128:104047 [FREE Full text] [doi: 10.1016/j.jbi.2022.104047] [Medline: 35257868]
23. Wang H, Liu M, Hong S, Zhuang Y. A historical review and bibliometric analysis of GPS research from 1991–2010. Scientometrics 2012 Sep 28;95(1):35-44. [doi: 10.1007/s11192-012-0853-z]
24. Guo L, Xu F, Feng Z, Zhang G. A bibliometric analysis of oyster research from 1991 to 2014. Aquacult Int 2015 Jul 10;24(1):327-344. [doi: 10.1007/s10499-015-9928-1]
25. Li X, Qian M, Xu B, Shi D, He P. Research on the construction strategy of cloud pharmacy of internet hospitals in Shanghai. China Digital Medicine 2021 Dec;16(12):29-33 (in Chinese) [FREE Full text] [doi: 10.3969/j.issn.1673-7571.2021.12.007]
26. Tian R. Comparative analysis of mobile internet appointment platforms in Beijing hospitals. Medical Education Management 2021;7(S1):217-220 (in Chinese) [FREE Full text]
27. Zhang R, Wang G. "Internet+" medical service supply: model comparison and optimal pathway. Health Economics Research 2022;39(03):32-37 [FREE Full text] [doi: 10.14055/j.cnki.33-1056/f.2022.03.008]
28. Guan D, Sun Z. Analysis of dynamic resistance factors construction emphases of telemedicine development in the new period. Chinese Hospitals 2020;24(04):56-58 (in Chinese) [FREE Full text] [doi: 10.19660/j.issn.1671-0592.2020.04.19]
29. Lu C, Li Y, Guo M, Li G. Discussion on constructing "Moblie Medical + Medicare" to further promote graded medical treatment model. Medical Information Work 2017 Dec;38(11):53-59 [FREE Full text] [doi: 10.3969/j.issn.1673-6036.2017.11.012]
30. Liu X, Ma L, Yang F, Zhu J, Ma M. Study on the normative development of internet hospitals in China. J Med Inform 2022;43(05):21-24 [FREE Full text] [doi: 10.3969/j.issn.1673-6036.2022.05.004]
31. Shen J. The situation, connotation, and reflection on the digital reform of medical security in China. China Economic & Trade Herald 2021 Oct;2021(22):47-51 [FREE Full text]
32. Yao F. Exploring the construction of a smart healthcare system from the perspective of the internet: based on the Solow Economic Growth Model. East China Science and Technology 2022 Dec;2022(12):93-95 (in Chinese) [FREE Full text] [doi: 10.3969/j.issn.1006-8465.2022.12.028]
33. Gu J, Li R, Sun Y, Zhang J, Huang H, Zheng B, et al. Research on the realization path and key strategies of the high quality development of internet medicine in China. Chinese Hospitals 2022;26(06):33-35 (in Chinese) [FREE Full text]
34. Jiang M, Liu Y, Ma H, Duan Q. Evaluation of the quality and efficiency of clinical medical treatment in case of public health emergencies by using"Internet+"contactless management system in emergency center. Ningxia Medical Journal 2022;44(06):524-526 [FREE Full text]
35. Wu D, Cui W, Yu G. Analysis on the operation strategy of internet hospitals in China. Chinese Hospitals 2021;25(10):79-80 (in Chinese) [FREE Full text]
36. Gao J. Exploration of the internet plus home-based comfort health care model for chronic diseases. China Hospital CEO 2022;18(11):86-89 (in Chinese) [FREE Full text]
37. Yang X, Wu S, Dong P, Wu S, Chen Z. Construction and practice of internet medical quality control system. Chinese Hospitals 2022;26(01):17-19 (in Chinese) [FREE Full text]
38. Lu Q. Internet plus healthcare officially integrates with medical insurance. China Health 2021 Jan;2021(01):105-106 (in Chinese) [FREE Full text]
39. Xu Z. China Management Informationization 2021;24(01):125-127 [FREE Full text]
40. Li T, Feng H. Digital health community empowers grassroots health governance transformation. Administration Reform 2022;2022(08):56-63 (in Chinese) [FREE Full text]
41. Liu W. Internet-based diagnosis treatment: strong regulatory oversight promotes transformation. China Hospital CEO 2021;17(20):22-25 (in Chinese) [FREE Full text]
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]
43. Li H. SWOT analysis of medical insurance management in tertiary hospitals under internet medical model. Chinese Medical Record 2021;22(11):32-33 (in Chinese) [FREE Full text] [doi: 10.3969/j.issn.1672-2566.2021.11.013]
44. Deng Y, Zhou Y. Regulatory deficiencies comprehensive governance in internet plus healthcare insurance services. China Hospital CEO 2021 May;17(07):66-69 (in Chinese) [FREE Full text]


## Abbreviations

**CNKI:** China National Knowledge Infrastructure
**IPHC:** Internet Plus Healthcare

**TLS:** total link strength

*Edited by T de Azevedo Cardoso; submitted 21.08.23; peer-reviewed by J Wang, D Singh; comments to author 02.02.24; revised version received 04.02.24; accepted 03.05.24; published 23.07.24.*

*Please cite as:*
*Hu Z, Qin X, Chen K, Huang YN, Wang RS, Tung TH, Chuang YC, Wang BL*
*Chinese Health Insurance in the Digital Era: Bibliometric Study*
*Interact J Med Res 2024;13:e52020*
*URL: https://www.i-jmr.org/2024/1/e52020*
*doi:10.2196/52020*
*PMID:*

Original Paper

# Antibiotic Prescribing Behavior of Physicians in Outpatient Departments in Hospitals in Northwest Ethiopia: Structural Equation Modeling Approach

Asrat Agalu Abejew[1,2*], BPharm, MSc; Gizachew Yismaw Wubetu[3*], BSc, MSc, PhD; Teferi Gedif Fenta[1*], BPharm, MPH, PhD

[1]Department of Pharmaceutics and Social Pharmacy, School of Pharmacy, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

[2]Department of Pharmacy, College of Medicine and Health Sciences, Bahir Dar University, Bahir Dar, Ethiopia

[3]Amhara Public Health Institute, Bahir Dar, Ethiopia

[*]all authors contributed equally

**Corresponding Author:**
Asrat Agalu Abejew, BPharm, MSc
Department of Pharmaceutics and Social Pharmacy, School of Pharmacy
College of Health Sciences
Addis Ababa University
Zambia Street, Tikur Anbessa Specialized Hospital Compound
Addis Ababa, 2QC2+23Q
Ethiopia
Phone: 251 0917156682
Fax: 251 1176
Email: asruphar@gmail.com

## Abstract

**Background:** Antibiotic resistance, fueled by irrational prescribing, is a global threat associated with health, social, and economic consequences. Understanding antibiotic prescribing behavior and associated factors is important to promote good prescribing practice.

**Objective:** This study aimed to determine the factors affecting antibiotic prescribing behaviors of physicians based on the theory of planned behavior in hospitals in northwest Ethiopia in 2022.

**Methods:** A cross-sectional study was conducted from September 2022 to October 2022. A total of 185 health professionals were included, and a self-administered questionnaire was used to collect data. A structural equation model based on the modified theory of planned behavior was used to determine factors affecting antibiotic prescribing behavior. The percentages of physicians' estimated prescriptions for patients with upper respiratory tract infections (URTIs) and during weekly outpatient visits were used to predict antibiotic prescribing behavior and finally linked with behavioral constructs. A *P* value <.05 was considered significant.

**Results:** Physicians estimated that they prescribed antibiotics for 54.8% (9896/18,049) of weekly outpatient encounters, and 178 (96.2%) of the 185 physicians estimated they prescribed antibiotics for patients who presented with symptoms of a URTI. Physicians aged ≤30 years were less likely to prescribe antibiotics (48/100, 48%) for patients who presented with a URTI than physicians older than 30 years (51/100, 51%; *P*=.004), and general practitioners were less likely to prescribe antibiotics (47/100, 47%) for patients who presented with a URTI than residents (51/100, 51%; *P*=.03). Similarly, during outpatient visits, physicians ≤30 years old were less likely to prescribe antibiotics (54/100, 54%) than physicians older than 30 years (57/100, 57%; *P*<.001), male physicians were less likely to prescribe antibiotics (53/100, 53%) than female physicians (64/100, 64%; *P*=.03), and general practitioners were less likely to prescribe antibiotics (53/100, 53%) than residents (57/100, 57%; *P*=.02). Physicians with good knowledge were less affected by perceived social pressure (mean 4.4, SD 0.6) than those with poor knowledge (mean 4.0, SD 0.9; *P*<.001) and felt it was easy to make rational decisions (mean 4.1, SD 1.1) compared with those with poor knowledge (mean 3.8, SD 1; *P*<.001). However, intentions to reduce and prescribe antibiotics were not affected by attitudes, subjective norms, or perceived behavioral control, and perceived antibiotic prescribing behavior was not related to intentions to reduce or prescribe antibiotics.

**Conclusions:** Antibiotic prescribing behavior was not under the volitional control of physicians. This calls for a systematic approach to change antibiotic prescribing practices in hospital.

*(Interact J Med Res 2024;13:e57285)* doi:10.2196/57285

**KEYWORDS**

antibiotic prescribing behavior; Ethiopia; outpatient departments; physicians; SEM; TPB

## Introduction

Antimicrobial resistance (AMR) is a natural phenomenon [1], to which overuse and misuse of antibiotics contribute and augment [1-4]. Globally, antibiotic consumption has increased (eg, by 65% from 2000 to 2015), and 30% to 50% of antibiotic prescriptions were used either inappropriately or unnecessarily [2], further resulting in increased inappropriate use [3,4] and the development of selective pressure on antibiotics [5-7]. Inappropriate prescribing is a key contributing factor to the emergence of AMR [1,8,9] and varies from 62.8% for respiratory tract infections to 78.5% in patients with skin and soft tissue infections [4]. This would strengthen the belief that antibiotics ought to be prescribed and are effective in circumstances when they are not [10]. Physicians' prescribing behaviors impact not only patient health but also medical expenses and health resources [11]. It is recommended to monitor antibiotic prescribing in hospitals to improve the quality of antibiotic prescribing through education and practice changes [12**].** Identifying key behaviors and drivers for the behaviors that may be amenable to change and improve prescribing decisions is an important component of interventions in health care practice to mitigate the burden of AMR [1,10,11]. Antibiotic stewardship programs (ASPs), which are among the most common interventions in health facilities to optimize antibiotic use, are effective, low-cost methods to change behaviors that drive excessive prescribing of antibiotics in health facilities [1].

Human behavior is guided by beliefs about the likely consequence of the behavior (behavioral beliefs), beliefs about the normative expectations of others (normative beliefs), beliefs about the presence of factors that may facilitate or impede the performance of the behavior (control beliefs), shaping attitudes, subjective norms (SNs), and perceived behavioral control (PBC) [13]. It is reported that these behavioral beliefs (attitudes, PBC, and SN) of physicians are predictors of indiscriminate antibiotic prescribing behaviors in hospitals [9,14]; thus, campaigns that address both health service personnel and the general population should take this into account [8]. A high level of knowledge is known to be associated with a more positive attitude and behavioral intention for reducing antibiotic prescriptions and was linked with less complacency, less fear, and less ignorance, although it had indirect effects on intentions to prescribe antibiotics through the attitude of ignorance [14]. On the other hand, perceived higher patient pressure negatively affects attitudes toward the rational use of antibiotics and promotes higher use of antibiotics [15]. Thus, characterizing and designing behavior change interventions based on the behavior change wheel model and theory of planned behavior (TPB) serve as a framework for modeling the antibiotic prescribing behaviors of physicians [13,16,17]. Optimizing antibiotic consumption and reducing the rate of AMR are currently global issues [1,18]. In low- and middle-income countries, the prescribing of antibiotics is highly influenced by inadequate diagnostic facilities, lack of guidelines, difficulty monitoring patient progress, poor intensive care facilities, patient demand for quick relief, perceived patient expectations from past prescriptions, and fear of losing patients to competition [19,20]. This results in high mortality and morbidity due to inadequate regulation, limited access to diagnostic facilities, and antimicrobial over-prescription [21,22]. Based on the behavior change wheel, once a problem is identified and context is considered, functions and policies may be implemented as interventions to understand and change prescribing behavior and improve antibiotic consumption [13,17]. This requires the design and implementation of sustained awareness campaigns to change behaviors and improve health outcomes [9].

In sub-Saharan Africa, physicians still prescribe antibiotics based only on a simple assessment of patients' symptoms, just as they used to when antibiotics first became commonly used in the 1950s [9], due to a lack of diagnostic and antibiotic susceptibility tests, resulting in up to 95% of antibiotic prescriptions as unnecessary [23]. Prescribing antibiotics requires balancing physician, patient, and facility-related factors [24]. In Ethiopia, antibiotic prescribing in hospitals may account for 52.39% of all prescriptions [25], and one-half of prescribed antibiotics might not be needed [26]. Although behavior change campaigns can be very cost-effective for changing antibiotic prescribing practices, based on identified gaps [9], in Ethiopia, to our knowledge, there have been no studies to model the antibiotic prescribing behavior of physicians other than determining the perceptions of health professionals on AMR and antibiotic use [27,28]. Modeling behavior is needed to help clinical leaders drive ASP and design educational programs to help standardize and improve antibiotic prescribing behaviors in health facilities [29]. Thus, this study assessed the determinants of antibiotic prescribing behavior among physicians serving in outpatient departments (OPDs) in hospitals in northwest Ethiopia using a structural equation modeling (SEM) approach.

## Methods

### Study Area and Period

A cross-sectional study was conducted from September 2022 to October 2022 in 4 hospitals: Felege Hiwot Comprehensive Specialized Hospital, Tibebe Ghion Specialized Hospital, Debre Markos Comprehensive Specialized Hospital, and Injibara General Hospital. Except for attempts to implement ASPs in inpatient wards in some of the hospitals, there is currently no system to monitor antibiotic prescribing or enabling factors for prescribing antibiotics in OPDs. This survey assessed the knowledge, attitudes, SN, and PBC of physicians and their intention to prescribe antibiotics as possible factors for antibiotic prescribing behaviors to provide insights into the driving forces of antibiotic prescribing as a complementary factor for antibiotic

consumption, which together constitute baseline information to design effective ASPs to tackle AMR.

### Study Participants, Sample Size Determination, and Sampling Procedures

Physicians (general practitioners and residents) working in OPDs of internal medicine, pediatrics, gynecology and obstetrics, and surgical departments were included in the study. The sample size for health professionals was determined based on the following formula for a finite population:

$$n = \chi^2 NP (1\text{-}P)/ (d^2 (N - 1) + \chi^2 P (1 - P))$$

where n is the sample size and $\chi^2$ is the table value of the chi square for 1 degree of freedom at the desired confidence level ($1.96 \times 1.96=3.84$), *N* is the total population, *P* is the population proportion (27%), and *d* is the degree of accuracy expressed as a proportion (0.05). According to Gebretekle et al [27], physicians estimate that they prescribe antibiotics to about 27% of their patients. Thus, a prevalence of 27% was used to calculate the sample size in this study. Accordingly, n was calculated as follows:

$$n=(1.96 \times 1.96) \times 487 \times 0.27(1–.27)/((0.05\times0.05) \times (487–1) + (1.96 \times 1.96) \times 0.27(1–0.27))$$

$$n=181.26 \text{ or } \sim182$$

To account for the nonresponse rate, 10% was added; thus, the total sample size was 200.

### Data Collection Instruments and Processes

Data collection was based on the study by Liu et al [14,15] and customized to local scenarios in Ethiopian hospital settings. Questionnaires consisted of 4 behavioral aspects leading to antibiotic prescribing based on the TPB, namely attitudes (the degree to which a prescriber is in favor of the use of antibiotics), SN (perceived social pressure to which a prescriber is subject to prescribe antibiotics), PBC (the ease or difficulty of making a rational decision on antibiotic prescriptions), and intentions (the degree to which a prescriber is willing to prescribe or reduce antibiotics). The questionnaire for professionals was designed on a Likert scale with a 5-point response format, ranging from 1 (strongly disagree) to 5 (strongly agree) for attitudes about and intention to prescribe antibiotics, and a 5-point response format (from always to never) for SN and PBC. In addition, physicians were asked to estimate the number of patients who receive antibiotics from their weekly encounters that involve prescriptions and the number of patients for whom they prescribe antibiotics from 10 encounters with patients with symptoms of upper respiratory tract infections (URTIs) to assess their antibiotic prescribing behavior or practices. To assess physicians' knowledge, 11 questions were used, attitude was assessed using 7 questions, SN was assessed using 8 questions, PBC was assessed using 5 questions, and there were 3 questions each to measure intentions to reduce and prescribe antibiotics.

Physicians (general practitioners and residents) working in internal medicine, pediatrics, gynecology and obstetrics, and surgical OPDs in the hospital were approached to participate in the study. The questionnaire was distributed while they were on duty. The completeness of the data was monitored on a daily basis. Finally, the data were compiled, and the behavioral constructs were linked with the percentages of physicians' perceived antibiotic prescribing behaviors and practices using SEM based on modified TPB (MTPB).

### The Theoretical Framework for Structural Equation Modeling

Attitude, SN, and PBC were shown to be related to appropriate sets of salient behavioral, normative, and control beliefs about a behavior. PBC, together with behavioral intention, can be used directly to predict behavioral achievement. Attitude is defined as the degree to which a prescriber is in favor of the use of antibiotics in outpatient encounters, whereas SN and PBC measure the perceived social pressure to which a prescriber is subject to prescribe antibiotics and the perceived ease or difficulty of making a rational decision during antibiotic prescriptions, respectively. A behavioral intention that is intermediate measures the degree to which a prescriber is willing to prescribe antibiotics [13]. Thus, the theoretical framework was adopted from the TPB model [13], and links between knowledge and attitude, SN, and PBC were explored. However, since the comparative fit indexes (CFIs) were low, knowledge was linked to SN and PBC in relation to antibiotic use, and attitude, SN, and PBC were linked to intentions to prescribe antibiotics and finally to behaviors influencing antibiotic prescribing.

### Statistical Analysis

Data were coded, entered, cleaned, and transferred to STATA version 14.0 (Stata Corp) for SEM analyses, but descriptive statistics were analyzed using SPSS version 23 (IBM Corp). ANOVA and chi-squared tests were performed to determine the difference in the mean measuring knowledge, attitudes, SN, PBC, and behavioral intentions of the participants according to age, gender, city, professional status, workplace, and duration of clinical practice. For knowledge, the percentage of respondents who answered correctly and the total number of correct answers per respondent were calculated. In addition, correct answers were coded as 1, and incorrect answers were coded as 0 for the SEM. Each attitude item was coded using a 5-point Likert scale (1=strongly agree, 5=strongly disagree), then recoded (–2=strongly disagree, 2=strongly agree), with a negative score indicating disagreements and a positive score indicating agreement with the average scores (ranging from –2 to 2). Intentions to reduce and prescribe antibiotics were coded similarly as the attitude measurements, with a negative score indicating refusal and a positive score indicating support for reducing antibiotic prescriptions (from –2 to 2). SN and PBC were measured from 1 to 5, with 1 indicating always and 5 indicating never, then recoded from 0 to 4, where 0 denotes never and 4 represents always. Behaviors around antibiotic prescriptions were measured using the percentage of antibiotic prescriptions for URTIs, per every 10 patients, and the percentage of antibiotic prescriptions among the estimated weekly visits.

Each variable was modeled separately to exclude factor loadings <0.3. Finally, SEM was applied to establish the associations between knowledge, attitudes, and practices. Standardized path coefficients with statistical significance ($P<.05$) were used. The maximum likelihood method was used to estimate the

parameters. The fitness of the data in the SEM model was assessed using model fitness indexes based on recommended level acceptances such as $P>\chi^2$ ($P>.05$), standardized root mean squared residual <0.09, and root mean squared error of approximation <0.08; Tucker-Lewis index >0.90; CFI>0.90; and coefficient of determination ≥0.7. In addition, descriptive analysis was used.

### Operational Definitions

We considered attitude to be the degree to which a participant had a positive or negative evaluation of indiscriminate antibiotic use. SNs were participants' beliefs about whether significant others would approve or disapprove of indiscriminate antibiotic use (ie, the perceived social pressure to which a prescriber is subject to prescribe antibiotics). PBC was the participant's beliefs regarding the ease or difficulty of making a rational decision about antibiotic prescriptions. Knowledge was considered participants' understanding and awareness regarding indiscriminate antibiotic use and AMR. The level of knowledge was determined based on the average score for all the questions (ie, physicians who answered at least or above the average score were considered to have good knowledge). Behavioral intentions around antibiotic prescriptions were the degree to which a prescriber was willing to prescribe antibiotics. Behaviors were documented as physicians' self-reported antibiotic prescribing behaviors.

### Ethical Considerations

Ethical approval was obtained from the College of Health Sciences (protocol code: 106/22/SoP) and the School of Pharmacy (protocol code: ERB/SOP/472/14/2022) of Addis Ababa University. A support letter to the hospitals was obtained from the Amhara Public Health Institute. During data collection, physicians' names were deidentified. All participants provided informed consent prior to participating in this study. The information obtained was kept confidential and used only for research purposes. Ethical issues like privacy and confidentiality were considered during data collection in order not to disclose information about people outside the research.

## Results

Among 200 planned respondents, 185 completed the questionnaires. Thus, the overall response rate was 92.5%.

### General Characteristics of Study Participants

The majority of physicians (153/185, 82.7%) were men, and their average age was 30.3 (3.9) years. Overall, physicians had been in their current roles an average of 3.0 (2.2) years and had worked at their current hospital for an average of 2.0 (1.8) years. The majority of physicians (149/185, 80.5%) had worked in their current roles for <5 years; 87 (87/185, 47%) and 55 (55/185, 29.9%) were from Tibebe Ghion Specialized Hospital and the gynecology and obstetrics department, respectively. Of the patients who visited the OPDs, the physicians estimated that 9896 (9896/18,049 54.8%) had received at least one antibiotic (Table 1).

**Table 1.** Characteristics of the 185 physician respondents in hospital outpatient departments in 2022.

| Variables | Results |
|---|---|
| **Sex, n (%)** | |
| Male | 153 (82.7) |
| Female | 32 (17.3) |
| **Age (years), n (%)** | |
| ≤30 | 116 (62.7) |
| >30 | 69 (37.3) |
| **Professional title, n (%)** | |
| GP[a] | 93 (50.2) |
| Residents | 92 (49.8) |
| **Facility, n (%)** | |
| FHCSH[b] | 43 (23.2) |
| TGSH[c] | 87 (47) |
| DMCSH[d] | 31(16.8) |
| IGH[e] | 24 (13) |
| **Department, n (%)** | |
| Medical | 36 (19.5) |
| Surgery | 40 (21.6) |
| Pediatrics | 48 (25.9) |
| Gyne/obs[f] | 55 (29.9) |
| Others | 6 (3.2) |
| **Length of employment in the current hospital (years), n (%)** | |
| <5 | 172 (93) |
| ≥5 | 13 (7) |
| **Overall length of employment in the current role (years), n (%)** | |
| <5 | 149 (80.5) |
| ≥5 | 36 (19.5) |
| **Training on antibiotic use** | |
| No | 159 (85.9) |
| Yes | 26 (14.1) |
| Patients seen per week, n | 18,049 |
| Patients seen per week, mean (SD) | 97.5 (79.7) |
| Patients estimated to receive antibiotics per week, n | 9896 |
| Patients estimated to receive antibiotics per week, mean (SD) | 60.1 (53.5) |

[a]GP: general practitioner.

[b]FHCSH: Felege Hiwot Comprehensive Specialized Hospital.

[c]TGSH: Tibebe Ghion Specialized Hospital.

[d] DMCSH: Debre Markos Comprehensive Specialized Hospital.

[e]IGH: Injibara General Hospital.

[f]Gyne/obs: gynecology and obstetrics.

### Knowledge of Physicians About Antibiotic Prescription

The majority of physicians agreed that amoxicillin is safe for pregnant patients (169/185, 91.4%), metronidazole has the best activity against anaerobes (166/185, 89.7%), and antibiotics should not be prescribed for nonfebrile diarrhea (151/185, 81.6%). However, none of the physicians answered, "Aminoglycosides are very active if they are administered parenterally once daily." Physicians answered 5 of 11 (46%) questions correctly. Based on this, 121 (65.4%) of the 185 physicians had good knowledge, based on the cutoff of a mean ≥5; however, for 64 (34%) of the 185 physicians, knowledge was poor (Table 2).

**Table 2.** Knowledge about antibiotic prescriptions of 185 physicians in hospital outpatient departments in 2022.

| Code | Items to assess knowledge levels | Response | | Factor loading | *P* value |
|---|---|---|---|---|---|
| | Correct, n (%) | Incorrect, n (%) | | | |
| q35 | Antibiotic treatment is not needed for non-febrile diarrhea. | 151 (81.6) | 34 (19.4) | 0.43 | <.001 |
| q36 | Antibiotics are not prescribed for upper respiratory tract infections. | 19 (10.3) | 166 (89.7) | –0.14 | .11 |
| q37 | Dosage reduction for ceftriaxone and clindamycin is needed for renal failure. | 37 (20) | 148 (80) | –0.11 | .22 |
| q38 | Amoxicillin is a safe antibiotic product for pregnant patients. | 169 (91.4) | 16 (9.9) | 0.55 | <.001 |
| q39 | Metronidazole has the best activity against anaerobes. | 166 (89.7) | 19 (10.3) | 0.82 | <.001 |
| q40 | Methicillin-resistant *staphylococcus aureus* is resistant to beta-lactam antibiotics. | 108 (58.4) | 77 (39.4) | 0.42 | <.001 |
| q41 | Ceftriaxone most effectively crosses the blood-brain barrier. | 85 (45.9) | 100 (54.1) | 0.37 | <.001 |
| q42 | Aminoglycosides are very active if they are administered parenterally once daily. | 0 | 185 (100) | 0 | .99 |
| q43 | Bacterial pneumonia (with symptoms of fast breathing, chest in-drawing, or stridor) requires antibiotic treatment. | 90 (48.6) | 95 (48.6) | 0.14 | .15 |
| q44 | Antibiotics do not reduce the duration and the occurrence of complications of upper respiratory tract infections. | 37 (20) | 148 (80) | 0.17 | .07 |
| q45 | The average number of patients taking antibiotics should be below 30% in a primary care facility. | 36 (19.5) | 149 (80.5) | 0.063 | .47 |

### Attitudes and Intentions of Physicians Toward Antibiotic Prescriptions

The mean response for attitude questions was 2.5 (0.4). Of the 185 physicians, 88 (47.6%) perceived that microbiology results are important for treating infectious diseases, 95 (51.4%) believed that over-prescribing of antibiotics contributes to the generation of antibiotic resistance, and 89 (48.1%) believed that over-prescription of antibiotics leads to the development of resistance. Regarding intention to prescribe antibiotics, the mean score for intention to reduce antibiotics was 2.4 (0.9), whereas the mean score for intention to prescribe antibiotics 2.5 (0.8). Of the 185 physicians, 133 (71.9%) wanted to reduce antibiotic consumption, 132 (71.4%) expected to reduce antibiotic consumption, and 117 (63.2%) planned to reduce antibiotic consumption for outpatients; however, 107 (57.8%) wanted to prescribe antibiotics, 103 (55.6%) expected to prescribe antibiotics, and 102 (55.1%) planned to prescribe antibiotics to their patients (Table 3).

**Table 3.** Physicians' (n=185) responses to individual items about attitudes and behavioral intentions about antibiotic prescribing in hospital outpatient departments in 2022.

| Measurement and items | Code | Response score, mean (SD) | Responses, n (%) | | | | | Factor loading | *P* value |
|---|---|---|---|---|---|---|---|---|---|
| | | | Strongly agree 1) | Agree (2) | Neutral (3) | Disagree (4) | Strongly disagree (5) | | |
| **Attitude**[a] | | | | | | | | | |
| In primary care, microbiology results are useful when treating infectious diseases. | Q1 | 1.8 (1) | 88 (47.6) | 73 (39.5) | 8 (4.3) | 11 (5.9) | 5 (2.7) | 0.34 | <.001 |
| The prescription of an antibiotic to a patient does not influence the development of resistance. | Q2 | 4.2 (0.9) | 0 | 16 (8.6) | 12 (6.5) | 68 (36.8) | 89 (48.1) | –0.09 | .30 |
| Overuse of antibiotics contributes to the generation of antibiotic resistance. | Q3 | 1.7 (1) | 107 (57.8) | 57 (30.8) | 7 (3.8) | 7 (3.8) | 7 (3.8) | 0.56 | <.001 |
| Prescribing antibiotics to patients does not cause damage even if they are not indicated. | Q4 | 4.2 (0.9) | 1 (0.01) | 13 (7.0) | 10 (5.4) | 77 (41.6) | 84 (45.4) | –0.03 | .73 |
| Over-prescribing antibiotics contributes to the generation of antibiotic resistance | Q5 | 1.7 (0.9) | 95 (51.4) | 64 (34.6) | 13 (7) | 10 (5.4) | 3 (1.6) | 0.80 | <.001 |
| Irrational use of broad-spectrum antibiotics contributes to generation of AMR[b] | Q6 | 1.7 (0.9) | 88 (47.6) | 81 (43.8) | 7 (3.8) | 5 (2.7) | 4 (2.2) | 0.62 | <.001 |
| Not selecting antibiotics to be prescribed based on the infected bacteria contributes to the generation of antibiotic resistance. | Q7 | 2.2 (1.1) | 56 (30.3) | 82 (44.3) | 16 (8.6) | 22 (11.9) | 9 (4.9) | 0.27 | .009 |
| **Intention to reduce antibiotics**[c] | | | | | | | | | |
| I want to reduce antibiotic consumption for outpatients. | Q29 | 2.5 (1.7) | 20 (10.8) | 113 (61.1) | 27 (14.6) | 18 (9.7) | 7 (3.8) | 0.40 | <.001 |
| I expect to reduce antibiotic consumption for outpatients. | Q30 | 2.3 (0.8) | 21 (11.4) | 111 (60) | 33 (17.8) | 18 (9.7) | 2 (1.1) | 0.82 | .001 |
| I plan to reduce antibiotic consumption for outpatients. | Q31 | 2.4 (0.8) | 15 (8.1) | 102 (55.1) | 48 (25.9) | 18 (9.7) | 2 (1.1) | 0.57 | <.001 |
| **Intention to prescribe antibiotics**[d] | | | | | | | | | |
| I want to prescribe antibiotics to outpatients | Q32 | 2.5 (0.9) | 23 (12.4) | 84 (45.4) | 55 (29.7) | 17 (9.2) | 6 (3.2) | 0.74 | <.001 |
| I expect to prescribe antibiotics to outpatients. | Q33 | 2.5 (0.9) | 16 (8.6) | 87 (47) | 61 (33) | 17 (9.2) | 4 (2.2) | 0.87 | <.001 |
| I plan to prescribe antibiotics to outpatients. | Q34 | 2.5 (0.9) | 15 (8.1) | 87 (47) | 57 (30.1) | 21 (11.4) | 5 (2.7) | 0.80 | <.001 |

[a]Overall score: mean 2.5 (SD 0.4).

[b]AMR: antimicrobial resistance.

[c]Overall score: mean 2.4 (SD 0.9).

[d]Overall score: mean 2.5 (SD 0.8).

## Subjective Norms and Perceived Behavioral Control of Physicians

The mean scores for SNs and PBC were 4.3 (0.8) and 4.0 (1.1), respectively. Of the 185 physicians, 133 (71.9%), 128 (69.2%), 126 (68.1%), and 125 (67.6%) never prescribed antibiotics based on patients' expectations, based on patient pressure, based on patients' requests for antibiotics, and to make patients trust them, respectively. Similarly, 119 (64.3%) of the 185 physicians never prescribed antibiotics to avoid being perceived as doing nothing for patients. Only a limited number of physicians agreed that they prescribed antibiotics based on patients' expectations or pressure (Table 4).

**Table 4.** Physicians' (n=185) responses to individual items about subjective norms and perceived behavioral control for intention to prescribe antibiotics in hospital outpatient departments in 2022.

| Measurement and items | Codes | Response score, mean (SD) | Responses, n (%) | | | | | Factor loading | *P* value |
|---|---|---|---|---|---|---|---|---|---|
| | | | Always (1) | Often (2) | Sometimes (3) | Rarely (4) | Never (5) | | |
| **Subjective norm**[a] | | | | | | | | | |
| I prescribe antibiotics since patients expect it. | Q15 | 3.9 (1.2) | 6 (3.2) | 21 (11.4) | 38 (20.5) | 48 (25.9) | 72 (38.9) | 0.54 | <.001 |
| I prescribe antibiotics since patients require and insist on it. | Q16 | 3.9 (1.1) | 7 (3.8) | 16 (8.6) | 33 (17.8) | 55 (29.7) | 74 (40) | 0.64 | <.001 |
| I prescribe antibiotics to satisfy patients. | Q17 | 4.1 (1.2) | 10 (5.7) | 12 (6.5) | 24 (13.0) | 46 (24.9) | 93 (50.3) | 0.55 | <.001 |
| I prescribe antibiotics so patients continue to trust me. | Q18 | 4.4 (1.0) | 5 (2.7) | 11 (5.9) | 12 (6.5) | 32 (17.3) | 125 (67.6) | 0.68 | <.001 |
| Even when I know that they are not indicated, I prescribe antibiotics since patients expect it. | Q19 | 4.5 (0.9) | 3 (1.6) | 4 (2.2) | 19 (10.3) | 26 (14.1) | 133 (71.9) | 0.84 | <.001 |
| Even when I know that they are not indicated, I prescribe antibiotics since patients ask for it. | Q20 | 4.5 (0.8) | 1 (0.1) | 7 (3.8) | 13 (7) | 38 (20.5) | 126 (68.1) | 0.86 | <.001 |
| Even when I know that they are not indicated, I prescribe antibiotics since patients press me to prescribe it. | Q21 | 4.5 (0.9) | 3 (1.6) | 3 (1.6) | 19 (10.3) | 32 (17.3) | 128 (69.2) | 0.87 | <.001 |
| Even when I know that they are not indicated, I prescribe antibiotics since I do not have time to explain to the patient the reason why they are not called for. | Q22 | 4.3 (1.1) | 4 (2.2) | 16 (8.6) | 10 (5.4) | 40 (21.6) | 115 (62.2) | 0.70 | <.001 |
| **Perceived behavioral control**[b] | | | | | | | | | |
| I prescribe antibiotics because I fear patient deterioration. | Q23 | 3.9 (3.2) | 5 (2.7) | 25 (13.5) | 50 (27) | 43 (23.2) | 62 (33.5) | 0.28 | <.001 |
| I prescribe antibiotics since it is impossible to track the patient accurately. | Q24 | 3.7 (1.1) | 4 (2.2) | 24 (13) | 47 (25.4) | 55 (29.7) | 55 (29.7) | 0.63 | <.001 |
| I prescribe antibiotics to avoid possible patient complaints or medico-legal problems. | Q25 | 4.1 (1.1) | 8 (4.3) | 10 (5.4) | 31 (16.8) | 44 (23.8) | 92 (49.7) | 0.85 | <.001 |
| I prescribe antibiotics to avoid being perceived as doing nothing for patients. | Q26 | 4.3 (1.1) | 6 (3.2) | 10 (5.4) | 18 (9.7) | 32 (17.3) | 119 (64.3) | 0.73 | <.001 |
| I prescribe antibiotics to avoid losing patients. | Q27 | 3.9 (1.3) | 9 (4.9) | 25 (13.5) | 23 (12.4) | 32 (17.3) | 96 (51.9) | 0.61 | <.001 |

[a]Overall score: mean 4.3 (SD 0.8).

[b]Overall score: mean 4.0 (SD 1.1).

## Antibiotic Prescribing Practices of Physicians

Of the 18,049 patients seen in the OPDs, 9896 (54.8%), or an average of 60.1 (53.5%) patients per week, were estimated to receive at least one antibiotic. Using an estimate of 10 patients for each of the 185 physicians, for a total of 1850 patients with URTIs, about 916 (49.5%) were estimated to be prescribed at least one antibiotic. Accordingly, 178 (96.2%) of the 185 physicians estimated that they prescribed antibiotics for at least one patient out of every 10 patients who presented with symptoms of a URTI, with a mean score of 5.9 (SD 2.2); 142 (142/185, 76.8%) physicians believed they would prescribe antibiotics for >3 patients; and 43 (43/185, 23.3%) physicians estimated they would prescribe antibiotics for 0 to 3 patients. The majority of physicians (56/185, 30.3%) said they would prescribe antibiotics to 5 patients out of 10 encounters with patients with URTIs in the OPDs (Table 5).

**Table 5.** Estimated prescriptions of antibiotics for upper respiratory tract infections (URTIs) out of every 10 patients by 185 physicians in hospital outpatient departments.

| Number of patients prescribed antibiotics per every 10 patients | Physicians who estimated they would prescribe antibiotics, n (%) |
|---|---|
| 0 | 7 (3.8) |
| 1 | 4 (2.2) |
| 2 | 9 (4.9) |
| 3 | 23 (12.4) |
| Total for ≤3 encounters (30% of patients) with a URTI | 43 (23.3) |
| 4 | 31 (16.8) |
| 5 | 56 (30.3) |
| 6 | 20 (10.8) |
| 7 | 10 (5.4) |
| 8 | 12 (6.5) |
| 9 | 2 (1.1) |
| 10 | 11 (5.9) |
| Total for >3 encounters with a URTI | 142 (76.8) |

## Structural Equation Modeling

The SEM using MTPB confirmed the theoretical framework for the antibiotic prescribing behaviors of physicians with some modifications. Based on the coefficient of determination ($R^2$), 94.6% of the variation in antibiotic prescribing behavior could be explained by all the exogenous variables. Data in the MTPB model had good fit, with $P>\chi^2$ ($P>0.0001$), a root mean squared error of approximation of 0.049, a standardized root mean squared residual of 0.071, a CFI of 0.91, and a Tucker-Lewis index of 0.901 (Table 6).

The MTPB model indicated that only physician knowledge was associated with PBC and SN. There was covariance between SNs and PBC ($P<.001$). Attitudes, SNs, and PBC were not associated with intentions to prescribe or reduce use of antibiotics. Similarly, intentions to prescribe or reduce use of antibiotics was not associated with the estimated number of antibiotic prescriptions for URTIs or during weekly visits (Figure 1). Physician age ($P=.004$) and professional level ($P<.02$) were predictors of the number of estimated prescriptions for URTIs, and physician age ($P=.001$), sex ($P=.03$), and professional level ($P=.02$) were predictors of the estimated number of prescriptions during weekly OPD visits. Knowledge was a direct predictor of SNs ($P<.001$) and PBC ($P<.001$). There was no indirect relationship between prescriber behaviors and knowledge, attitude, SN, and PBC (Table 7).

Based on the information in Table 7, for the 49.5% of the 1850 patients with URTIs who were estimated to be prescribed at least one antibiotic, physicians older than 30 years were more likely to prescribe antibiotics (51/100, 51%) than those ≤30 years old (48/100, 48%). Based on professional level, residents (51/100, 51%) were more likely to prescribe antibiotics than general practitioners (47/100, 47%). Similarly, for the estimated 54.8% (9896/18,049) of weekly OPD visits that had an antibiotic prescription, physicians older than 30 years were more likely to prescribe antibiotics (57/100, 57%) than those ≤30 years old (54/100, 54%). Women (63/100, 63%) and residents (57/100, 57%) were also more likely to prescribe antibiotics than men (53/100, 53%) and general practitioners (53/100, 53%), respectively. Good knowledge was a direct predictor of SNs (mean 4.4, SD 0.6) and PBC (mean 4.1, SD 1.1), both of which are in contrast for those with poor knowledge (mean 4.0, SD 0.9) and (mean 3.8, SD 1), respectively. However, intentions to reduce and prescribe antibiotics were not affected by attitudes, SNs, nor PBC, and perceived antibiotic prescribing behavior was not related to intentions to reduce or prescribe antibiotics.

**Table 6.** The model goodness of fit indexes for antibiotic prescribing behaviors of 185 physicians in hospital outpatient departments.

| Fit statistics | Value | Description | Standard |
|---|---|---|---|
| **Likelihood ratio** | | | |
| $\chi^2$_ms (314) | 451.292 | Model versus saturated | —[a] |
| $P>\chi^2$ | 0.0001 | — | — |
| $\chi^2$_bs (348) | 1878.872 | Baseline versus saturated | — |
| $P>\chi^2$ | 0.0001 | — | — |
| **Information criteria** | | | |
| Akaike information criterion (AIC) | 14073.713 | — | |
| Bayesian information criterion (BIC) | 14337.782 | — | |
| **Population error** | | | |
| Root mean squared error of approximation (RMSEA) | 0.049 | — | 0.08 |
| 90% CI | 0.038-0.058 | — | — |
| Pclose | 0.581 | Probability of RMSEA ≤0.05 | — |
| **Size of residuals** | | | |
| Standardized root mean squared residual (SRMR) | 0.071 | — | <0.09 |
| Coefficient of determination (CD) | 0.946 | — | — |
| **Baseline comparison** | | | |
| Comparative fit index (CFI) | 0.910 | — | ≥0.90 |
| Tucker-Lewis index (TLI) | 0.901 | — | ≥0.90 |

[a]Not applicable.

**Figure 1.** SEM of Antibiotic Prescribing Behavior in hospital outpatient departments using standardized coefficients in 2022. This SEM is also available as Multimedia Appendix 1. SEM: Structural Equation Modeling.

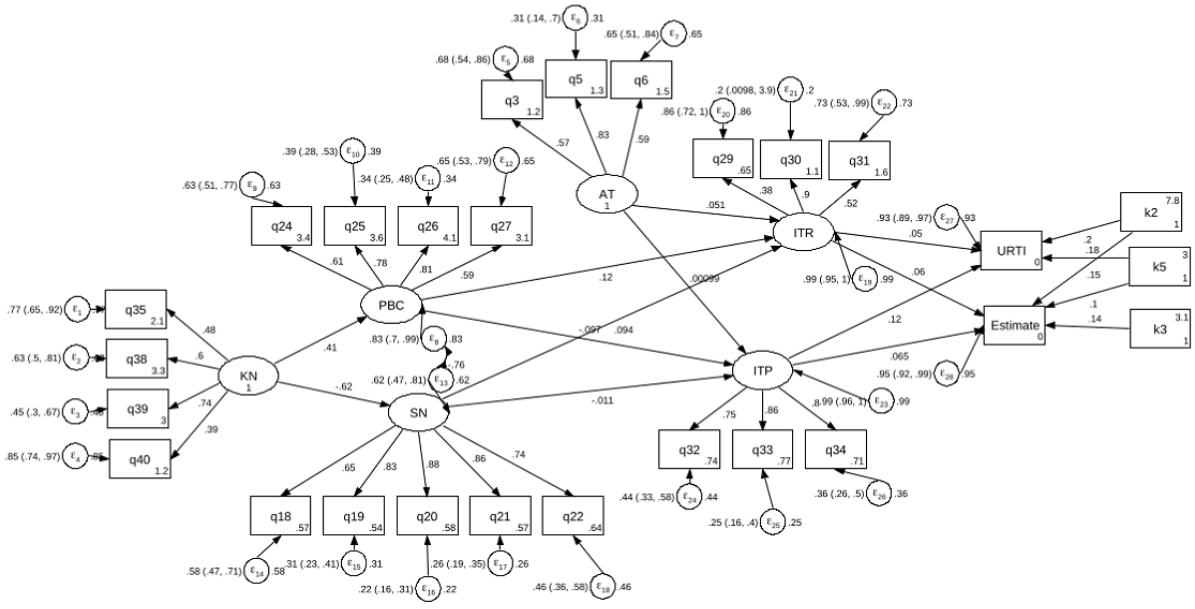


**Table 7.** Effects of direct and indirect variables on the intention to prescribe antibiotics by 185 physicians in hospital outpatient departments in 2022.

| Factor | Effects of covariates on each other | | |
|---|---|---|---|
| | B | SE | *P* value |
| **Subjective norms** | | | |
| Knowledge | 0.41 | 0.49 | <.001 |
| **Perceived behavioral control** | | | |
| Knowledge | –0.62 | 0.43 | <.001 |
| **Intention to reduce antibiotic prescriptions** | | | |
| Attitude | 0.16 | 0.20 | .42 |
| Subjective norms | 0.15 | 0.20 | .45 |
| Perceived behavioral control | 0.13 | 0.21 | .53 |
| **Intention to prescribe antibiotics** | | | |
| Attitude | 0.10 | 0.89 | .91 |
| Subjective norms | 0.20 | 0.93 | .83 |
| Perceived behavioral control | –0.26 | 0.89 | .79 |
| **Proportion of patients from 10 who were estimated to be prescribed antibiotics for a URTI**[a] | | | |
| Facility | 3.09 | 1.84 | .09 |
| Age | 1.05 | 0.37 | .004 |
| Sex | 5.19 | 4.09 | .20 |
| Work area | 0.17 | 1.56 | .91 |
| Professional level | 6.96 | 3.24 | .03 |
| Length of employment in the current hospital | 0.36 | 1.64 | .83 |
| Overall length of employment in the current position | –0.31 | 1.31 | .81 |
| Training | 3.68 | 4.28 | .39 |
| ITRABx[b] | 1.66 | 2.69 | .54 |
| ITPABx[c] | Constrained | Constrained | Constrained |
| **Proportion of weekly visits during which physicians estimated they would prescribe antibiotics** | | | |
| Facility | 1.19 | 2.31 | .61 |
| Age | 1.53 | 0.46 | .001 |
| Sex | 11.15 | 5.16 | .03 |
| Work area | –2.83 | 1.97 | .15 |
| Professional level | 9.69 | 4.08 | .02 |
| Length of employment in the current hospital | 1.97 | 2.07 | .34 |
| Overall length of employment in the current position | –1.03 | 1.65 | .53 |
| Training | –8.13 | 5.39 | .13 |
| ITRABx | 2.88 | 3.38 | .39 |
| ITPABx | 0.65 | 0.78 | .41 |

[a]URTI: upper respiratory tract infection.

[b]ITRABx: intention to reduce antibioitcs.

[c]ITPABx: intention to prescribe antibiotics.

## Discussion

AMR is a global crisis [27,29-32], calling for urgent action to resolve it. One of the important strategies for combating AMR is improving antibiotic prescribing practices by determining factors that affect physicians' prescribing behaviors [10,11,14]. Physicians' antibiotic prescribing practices are influenced by physician-related factors (knowledge, expertise, specific

prescription patterns, time constraints, and communication with patients), patient-related factors (preferences, expectations, knowledge, culture, economic status, and previous experience), and health system–related factors (guidelines, policies, regulations, and financial incentives) [24]. Thus, the TPB in its original or modified (MTPB) form provides a theoretical framework to identify the determinants of physicians' antibiotic prescribing behaviors [14,15,33]. This study, using SEM, confirmed the theoretical framework for physicians' antibiotic prescribing behaviors with some modifications. The MTPB model revealed that physician knowledge was associated with PBC and SN ($P$<.001). However, attitudes, SNs, and PBC did not influence intentions around prescriptions and perceived prescribing behaviors of physicians. Overall, the study revealed that physicians' perceived antibiotic prescribing behaviors were not affected by intentions to reduce and prescribe antibiotics.

A qualitative study in Ethiopia uncovered that "junior physicians (interns and residents) are more likely to prescribe broad-spectrum antibiotics, and they further speculate these practices are driven by poor knowledge" [27]. In this, 121 (65.4%) of the 185 physicians had good knowledge (answered 5 questions out of 11), although the respondents answered only 44.13% of the total questions about antibiotic prescriptions correctly. This was relatively low compared with the 55% to 86% for physicians in hospitals in China [14,15], Lao People's Democratic Republic, Democratic Republic of Congo, and Peru [34-36]. In this study, MTPB knowledge was directly linked with the PBC and SNs of physicians ($P$<.05). This was different than the findings of a study in China that reported a link between high knowledge and positive attitudes toward the rational use of antibiotics [14,15] but similar to a study that reported a link between high knowledge and decreased SNs to prescribe antibiotics [14]. This study revealed a lack of indirect links between knowledge and antibiotic prescribing behavior, which was similar to a study in China that reported a lack of a link between knowledge and antibiotic prescribing practices [14,15]. Knowledge helps physicians weigh the treatment options and increases the accuracy of risk perception; thus, physicians with different professional titles (such as resident physicians and general practitioners) and length of practice may have different attitudes toward antibiotics, based on findings in previous studies [37,38]. Although training can contribute to knowledge acquisition, this study revealed a lack of difference in knowledge between those who completed training on antibiotic prescribing and AMR and those who did not.

This study also revealed a lack of a link between attitudes, SNs, and PBC to prescribe antibiotics. This was different from a report in China that confirmed that "intentions to prescribe antibiotics are predicted by the attitudes, subjective norms, and sense of behavioral control of the prescribers" [33], although another study reported a lack of relationship with intentions to prescribe antibiotics [14,15]. It was reported that attitudes, SNs, and PBC were predictors of antibiotic prescribing behaviors [11,39]. However, this study did not show indirect relationships between attitudes, SNs, and PBC and prescribers' perceived antibiotic prescribing behaviors.

Of the patients seen in the OPDs weekly, 54.8% were estimated to receive at least one antibiotic, and of the 1850 estimated patients who presented with a URTI, 916 (49.5%) were estimated to be prescribed at least one antibiotic. In Ethiopia, antibiotic prescribing ranges from 56.0% in primary health care facilities [40] to 73.7% in inpatient wards in the national referral hospital [41]. The majority of the prescriptions (32.9%-39.3%) were for respiratory tract infections, although about 54.2% of all antibiotic prescriptions might not be needed [27]. This could explain the high rates of perceived antibiotic prescriptions for outpatient visits and patients with URTIs in this study. In ambulatory care facilities in Tanzania, 95% to 96.3% of presenting cases were receiving at least one antibiotic [42], which is higher than the estimated prescriptions among weekly visits and patients with URTIs in this study. Another study in the same hospital reported 66.9% of patients were treated with or prescribed at least one antibiotic among weekly visits [28], which was higher than the perceived weekly prescriptions and prescriptions for URTIs in this study. A study in China reported that physicians prescribed antibiotics to an estimated 40% of patients with URTIs [14], and actual antibiotic prescribing behavior was 44.3% [33], both of which are lower than the rates in this study.

In this study, intention to reduce antibiotic use and intention to prescribe antibiotics were not linked with perceived antibiotic prescribing behaviors. Similar findings from China supported the limited role of intentions on antibiotic prescribing behavior [33]. Intraphysician prescription variability is affected by the availability of clinical guidelines, experience, peer prescribing practices, pharmaceutical pressure, time pressure, financial considerations, individual practice patterns, practice volume, and relationships with patients [24]. Differences in these factors might explain the discrepancy in the findings from this study and those from a study in China [33], which reported that a positive attitude toward antibiotics resulted in a higher intention to prescribe antibiotics. The difference might be due to differences in the type of patients, clinical practice, and availability of structural and process controls in antibiotic prescribing. Two studies, one in Eritrea and one in Ethiopia, confirmed this by reporting that patient age (<18 years), gender (male), and the number of drugs in a prescription (≥2) were associated with the prescription of antibiotics [43,44]. Thus, physicians in different health systems may be subjected to different working environments and social pressures, which could affect their overall intention to prescribe and prescribing behavior.

In general, this study uncovered a lack of volitional control among physicians when prescribing antibiotics. This indicates the complex nature of antibiotic prescribing practices, which are influenced by various factors [24]. Thus, a campaign is needed to reduce over-prescriptions of antibiotics using a systems approach to addressing gaps in the knowledge and attitudes of prescribers [33]. Introducing educational programs and training on antibiotic prescribing practices and antibiotic resistance, preparing targeted guidelines to address gaps in antibiotic prescriptions for URTIs, and involving specialists will help address the gaps in antibiotic prescribing [24,33,45]. Providing antibiotics as a universal therapy due to gaps in knowledge and skills and financial or reputational incentives on the one hand and a lack of antibiotic and poor facility

regulations, the absence of a regulatory framework, and poor implementation of existing policies on the other hand might be drivers of inappropriate antibiotic prescriptions [46]. Thus, disease-specific prescribing guidelines like those for URTIs can facilitate the translation of intentions into practices, since the inability to make a clear diagnosis and over-prescription of antibiotics may be linked with physicians' clinical capacity rather than behavioral control intentions [24]. Despite this, self-regulation, outcome expectation, and anticipation of possible barriers may still have a considerable effect on prescribing behaviors and will help reduce inappropriate antibiotic use, since inappropriate prescribing of antibiotics in ambulatory care is known to be linked to current knowledge on antibiotics, ASPs, and AMR among prescribers [42].

Therefore, ASPs must have fiduciary responsibility for all health care institutions across the continuum of care [47]. A comprehensive approach through a hospital policy on the rational use of antibiotics is essential to developing and implementing an evidence-based antibiotic use policy and standard treatment guidelines for common infectious diseases, improving antimicrobial prescribing through educational and administrative means, and monitoring and providing feedback regarding antibiotic resistance, all of which are strategic approaches [48]. Knowledge of determinants that influence antibiotic prescription behavior is essential for the successful implementation of antimicrobial stewardship interventions [49]. In Ethiopia, at present, there are limited or no national or coordinated legislative or regulatory mandates designed to optimize the use of antimicrobial therapy through ASPs [6,26]. This research on behavioral determinants may have a substantial impact on designing policies on antibiotic prescribing behaviors and implementing effective, efficient, and evidence-based interventions. It urges strengthening efforts to improve prevention and control efforts for infectious diseases, including the adoption of ASPs in all health care facilities. Research is also needed to define the optimal elements and goals of ASPs in different health care settings; expand educational efforts on ASP; devise novel mechanisms to prevent the over-prescription of antibiotics; and implement rapid, point-of-care diagnostic tests that would enable appropriate prescription and care.

Overall, this study will help understand prescribing behaviors by applying the TPB model and will be helpful for regulating prescribing behaviors, improving clinical management, promoting physician-patient communication, and establishing a harmonious physician-patient relationship to improve rational prescribing behaviors in delivering high-quality medical services. Policymakers should also consider multiple scenarios rather than merely concentrating on creating awareness. The model can hopefully be incorporated as part of multilevel interventions designed to decrease irrational prescriptions for actual patients. Furthermore, it might help initiate a nationwide survey on factors affecting antibiotic prescribing behaviors, and more research is needed to explore the views of other stakeholders on antibiotic use. Although this study presents opportunities for future studies in the country, it does have its limitations. The sample size was relatively small, and self-administered questionnaires may not provide the possibility for respondents to verify their answers, resulting in socially desirable answers. Due to a lack of records for antibiotic prescriptions, it was not possible to determine the actual antibiotic prescribing behaviors of physicians. Another limitation might be that it did not include all comprehensive physician-related factors (expertise, knowledge, specific prescription patterns, time constraints, and communication with patients), patient-related factors (knowledge, preferences, expectations, culture, economic status, and previous experience), and health system–related factors (financial incentives, guidelines, policies, and regulations) that influence antibiotic prescribing practices. Thus, further studies using TPB with a large number of physicians are warranted.

In conclusion, there was a high level of estimated antibiotic prescriptions for URTIs and weekly outpatient visits. However, the perceived behaviors around antibiotic prescription were not affected by the intention to prescribe antibiotics or the intention to reduce antibiotic use. Although the physicians had a good level of knowledge about antibiotics, antibiotic resistance, and prescriptions, which were linked with the attitudes and SNs of physicians, intentions to reduce use and prescribe antibiotics were not significantly associated with attitudes, SNs, or PBC. This may show the complex nature of antibiotic prescriptions, which cannot be justified simply by intentions and behaviors of physicians, as determined based on TPB.

## Acknowledgments

We acknowledge Addis Ababa University, the office of graduate studies, and Bahir Dar University. We are also thankful to the hospitals and health professionals for giving us information. We are also thankful to JMIR Publications for providing article processing fee (APF) support for publication of this paper.

## Authors' Contributions

AAA was involved in study conception, study design, study execution, acquisition of data, data analysis and interpretation, manuscript writing, and manuscript review. TGF was involved in study conception, study design, supervision, manuscript writing, and manuscript review. GYW was involved in supervision, manuscript writing, and manuscript review.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Structural equation modelling file.
[ZIP File (Zip Archive), 74 KB - ijmr_v13i1e57285_app1.zip ]

## References


1. Jonas OB, Irwin AB, Franck CJ, Le Gall FG, Marquez PV. Drug-resistant infections: a threat to our economic future (Vol. 2) : final report (English). HNP/Agriculture Global Antimicrobial Resistance Initiative. Washington, D.C: World Bank Group URL: http://documents.worldbank.org/curated/en/323311493396993758/final-report [accessed 2024-10-05]
2. Klein EY, Van Boeckel TP, Martinez EM, Pant S, Gandra S, Levin SA, et al. Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. Proc Natl Acad Sci U S A 2018 Apr 10;115(15):E3463-E3470 [FREE Full text] [doi: 10.1073/pnas.1717295115] [Medline: 29581252]
3. Dellit TH, Owens RC, McGowan JE, Gerding DN, Weinstein RA, Burke JP, Infectious Diseases Society of America, Society for Healthcare Epidemiology of America. Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America guidelines for developing an institutional program to enhance antimicrobial stewardship. Clin Infect Dis 2007 Jan 15;44(2):159-177. [doi: 10.1086/510393] [Medline: 17173212]
4. Saleem Z, Saeed H, Hassali MA, Godman B, Asif U, Yousaf M, et al. Pattern of inappropriate antibiotic use among hospitalized patients in Pakistan: a longitudinal surveillance and implications. Antimicrob Resist Infect Control 2019 Nov 21;8:188 [FREE Full text] [doi: 10.1186/s13756-019-0649-5] [Medline: 31768252]
5. Sedláková MH, Urbánek K, Vojtov V, Suchánková H, Imwensi P, Kolář M. Antibiotic consumption and its influence on the resistance in Enterobacteriaceae. BMC Res Notes 2014 Jul 16;7:454 [FREE Full text] [doi: 10.1186/1756-0500-7-454] [Medline: 25027417]
6. Guidelines: A Practical Guide to Antimicrobial Stewardship in Ethiopian Hospitals. USAID Global Health Supply Chain Program. 2018 May. URL: https://www.ghsupplychain.org/practical-guide-antimicrobial-stewardship [accessed 2024-10-05]
7. Lee H, Loh Y, Lee J, Liu C, Chu C. Antimicrobial consumption and resistance in five Gram-negative bacterial species in a hospital from 2003 to 2011. Journal of Microbiology, Immunology and Infection 2015 Dec;48(6):647-654. [doi: 10.1016/j.jmii.2014.04.009]
8. Drug-resistant infections: a threat to our economic future. The World Bank. 2017 Mar. URL: https://tinyurl.com/4h7t3jf3 [accessed 2024-10-08]
9. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. The Review on Antimicrobial Resistance. 2014. URL: https://tinyurl.com/2s48wx2h [accessed 2024-10-05]
10. Pinder R, Sallis A, Berry D, Chadborn T. Behaviour change and antibiotic prescribing in healthcare settings Literature review and behavioural analysis. Public Health England. 2015 Feb. URL: https://tinyurl.com/8hu88vjv [accessed 2024-10-05]
11. Pan L, Zhao R, Zhao N, Wei L, Wu Y, Fan H. Determinants associated with doctors' prescribing behaviors in public hospitals in China. Ann N Y Acad Sci 2022 Jan;1507(1):99-107. [doi: 10.1111/nyas.14677] [Medline: 34476819]
12. Versporten A, Zarb P, Caniaux I, Gros M, Drapier N, Miller M, et al. Antimicrobial consumption and resistance in adult hospital inpatients in 53 countries: results of an internet-based global point prevalence survey. The Lancet Global Health 2018 Jun;6(6):e619-e629. [doi: 10.1016/S2214-109X(18)30186-4]
13. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
14. Liu C, Liu C, Wang D, Zhang X. Knowledge, attitudes and intentions to prescribe antibiotics: a structural equation modeling study of primary care institutions in Hubei, China. IJERPH 2019 Jul 05;16(13):2385. [doi: 10.3390/ijerph16132385]
15. Liu C, Liu C, Wang D, Zhang X. Intrinsic and external determinants of antibiotic prescribing: a multi-level path analysis of primary care prescriptions in Hubei, China. Antimicrob Resist Infect Control 2019;8:132 [FREE Full text] [doi: 10.1186/s13756-019-0592-5] [Medline: 31406571]
16. Kpokiri EE, Taylor DG, Smith FJ. Development of antimicrobial stewardship programmes in low and middle-income countries: a mixed-methods study in Nigerian hospitals. Antibiotics 2020 Apr 23;9(4):204. [doi: 10.3390/antibiotics9040204]
17. Wilkinson A, Ebata A, MacGregor H. Interventions to reduce antibiotic prescribing in LMICs: a scoping review of evidence from human and animal health systems. Antibiotics 2018 Dec 22;8(1):2. [doi: 10.3390/antibiotics8010002]
18. Michie S, van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implementation Sci 2011 Apr 23;6(1):42. [doi: 10.1186/1748-5908-6-42]
19. Sartelli M, C. Hardcastle T, Catena F, Chichom-Mefire A, Coccolini F, Dhingra S, et al. Antibiotic use in low and middle-income countries and the challenges of antimicrobial resistance in surgery. Antibiotics 2020 Aug 09;9(8):497. [doi: 10.3390/antibiotics9080497]
20. Rawson TM, Moore LSP, Tivey AM, Tsao A, Gilchrist M, Charani E, et al. Behaviour change interventions to influence antimicrobial prescribing: a cross-sectional analysis of reports from UK state-of-the-art scientific conferences. Antimicrob Resist Infect Control 2017 Jan 13;6(1):a [FREE Full text] [doi: 10.1186/s13756-017-0170-7]
21. Sulis G, Sayood S, Gandra S. Antimicrobial resistance in low- and middle-income countries: current status and future directions. Expert Rev Anti Infect Ther 2022 Feb;20(2):147-160. [doi: 10.1080/14787210.2021.1951705] [Medline: 34225545]

22. Hashmi MZ, Strezov V. Emerging Contaminants and Associated Treatment Technologies. Cham, Switzerland: Springer; 2020.
23. Wojcik G, Ring N, McCulloch C, Willis DS, Williams B, Kydonaki K. Understanding the complexities of antibiotic prescribing behaviour in acute hospitals: a systematic review and meta-ethnography. Arch Public Health 2021 Jul 23;79(1):134 [FREE Full text] [doi: 10.1186/s13690-021-00624-1] [Medline: 34301332]
24. Kasse GE, Humphries J, Cosh SM, Islam MS. Factors contributing to the variation in antibiotic prescribing among primary health care physicians: a systematic review. BMC Prim Care 2024 Jan 02;25(1):8 [FREE Full text] [doi: 10.1186/s12875-023-02223-1] [Medline: 38166736]
25. Sisay M, Gashaw T, Amare F, Tesfa T, Baye Y. Hospital-level antibacterial prescribing and its completeness in Ethiopia: did it adhere to good prescribing practice? IJGM 2020 Nov;Volume 13:1025-1034. [doi: 10.2147/ijgm.s280696]
26. Gebretekle GB, Haile Mariam D, Workeabeba AT, Fentie AM, Wondwossen AW, Alemayehu T, et al. Half of prescribed antibiotics are not needed: a pharmacist-led antimicrobial stewardship intervention and clinical outcomes in a referral hospital in Ethiopia. Front. Public Health 2020 Apr 9;8:1. [doi: 10.3389/fpubh.2020.00109]
27. Gebretekle GB, Haile Mariam D, Abebe W, Amogne W, Tenna A, Fenta TG, et al. Opportunities and barriers to implementing antibiotic stewardship in low and middle-income countries: Lessons from a mixed-methods study in a tertiary care hospital in Ethiopia. PLoS ONE 2018 Dec 20;13(12):e0208447 [FREE Full text] [doi: 10.1371/journal.pone.0208447]
28. Abejew AA, Wubetu GY, Fenta TG. Assessment of challenges and opportunities in antibiotic stewardship program implementation in Northwest Ethiopia. Heliyon 2024 Jun;10(11):e32663 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e32663]
29. Worldwide Antimicrobial Resistance National/International Network Group (WARNING) Collaborators. Ten golden rules for optimal antibiotic use in hospital settings: the WARNING call to action. World J Emerg Surg 2023 Oct 16;18(1):50 [FREE Full text] [doi: 10.1186/s13017-023-00518-3] [Medline: 37845673]
30. CDC. Antibiotic Resistance Threats in the United States 2019. Centers for Disease Control and Prevention. Atlanta, GA: U.S. Department of Health and Human Services CDC; 2019. URL: https://tinyurl.com/p8rzt9fa [accessed 2024-10-05]
31. Davey P, Marwick CA, Scott CL, Charani E, McNeil K, Brown E, et al. Interventions to improve antibiotic prescribing practices for hospital inpatients. Cochrane Database Syst Rev 2017 Feb 09;2(2):CD003543 [FREE Full text] [doi: 10.1002/14651858.CD003543.pub4] [Medline: 28178770]
32. Lee Y, Lu M, Shao P, Lu P, Chen Y, Cheng S, et al. Nationwide surveillance of antimicrobial resistance among clinically important Gram-negative bacteria, with an emphasis on carbapenems and colistin: Results from the Surveillance of Multicenter Antimicrobial Resistance in Taiwan (SMART) in 2018. Int J Antimicrob Agents 2019 Sep;54(3):318-328. [doi: 10.1016/j.ijantimicag.2019.06.009] [Medline: 31202925]
33. Liu C, Liu C, Wang D, Deng Z, Tang Y, Zhang X. Determinants of antibiotic prescribing behaviors of primary care physicians in Hubei of China: a structural equation model based on the theory of planned behavior. Antimicrob Resist Infect Control 2019 Jan 30;8(1):a [FREE Full text] [doi: 10.1186/s13756-019-0478-6]
34. Quet F, Vlieghe E, Leyer C, Buisson Y, Newton PN, Naphayvong P, et al. Antibiotic prescription behaviours in Lao People's Democratic Republic: a knowledge, attitude and practice survey. Bull. World Health Organ 2015 Mar 03;93(4):219-227. [doi: 10.2471/blt.14.142844]
35. Thriemer K, Katuala Y, Batoko B, Alworonga J, Devlieger H, Van Geet C, et al. Antibiotic prescribing in DR Congo: a knowledge, attitude and practice survey among medical doctors and students. PLoS ONE 2013 Feb 18;8(2):e55495. [doi: 10.1371/journal.pone.0055495]
36. García C, Llamocca LP, García K, Jiménez A, Samalvides F, Gotuzzo E, et al. Knowledge, attitudes and practice survey about antimicrobial resistance and prescribing among physicians in a hospital setting in Lima, Peru. BMC Clin Pharmacol 2011 Nov 15;11(1):18. [doi: 10.1186/1472-6904-11-18]
37. Alameddine M, AlGurg R, Otaki F, Alsheikh-Ali AA. Physicians' perspective on shared decision-making in Dubai: a cross-sectional study. Hum Resour Health 2020 May 07;18(1):33. [doi: 10.1186/s12960-020-00475-x]
38. Oerlemans AJ, Knippenberg ML, Olthuis GJ. Learning shared decision-making in clinical practice. Patient Education and Counseling 2021 May;104(5):1206-1212. [doi: 10.1016/j.pec.2020.09.034]
39. Byrne M, Miellet S, McGlinn A, Fish J, Meedya S, Reynolds N, et al. The drivers of antibiotic use and misuse: the development and investigation of a theory driven community measure. BMC Public Health 2019 Oct 30;19(1):1425 [FREE Full text] [doi: 10.1186/s12889-019-7796-8]
40. Worku F, Tewahido D. Retrospective assessment of antibiotics prescribing at public primary healthcare facilities in Addis Ababa, Ethiopia. Interdiscip Perspect Infect Dis 2018;2018:4323769-4323769 [FREE Full text] [doi: 10.1155/2018/4323769] [Medline: 29681933]
41. Gutema G, Håkonsen H, Engidawork E, Toverud E. Multiple challenges of antibiotic use in a large hospital in Ethiopia – a ward-specific study showing high rates of hospital-acquired infections and ineffective prophylaxis. BMC Health Serv Res 2018 May 3;18(1):326 [FREE Full text] [doi: 10.1186/s12913-018-3107-9]
42. Massele A, Rogers AM, Gabriel D, Mayanda A, Magoma S, Cook A, et al. A narrative review of recent antibiotic prescribing practices in ambulatory care in Tanzania: findings and implications. Medicina 2023 Dec 18;59(12):2195. [doi: 10.3390/medicina59122195]

43. Dereje B, Workneh A, Megersa A, Yibabie S. Prescribing pattern and associated factors in community pharmacies: a cross-sectional study using AWaRe classification and WHO antibiotic prescribing indicators in Dire Dawa, Ethiopia. Drugs - Real World Outcomes 2023 Jun 10;10(3):459-469 [FREE Full text] [doi: 10.1007/s40801-023-00367-1]
44. Amaha ND, Weldemariam DG, Abdu N, Tesfamariam EH. Prescribing practices using WHO prescribing indicators and factors associated with antibiotic prescribing in six community pharmacies in Asmara, Eritrea: a cross-sectional study. Antimicrob Resist Infect Control 2019 Oct 22;8(1):163 [FREE Full text] [doi: 10.1186/s13756-019-0620-5]
45. Labi A, Obeng-Nkrumah N, Bjerrum S, Aryee NAA, Ofori-Adjei YA, Yawson AE, et al. Physicians' knowledge, attitudes, and perceptions concerning antibiotic resistance: a survey in a Ghanaian tertiary care hospital. BMC Health Serv Res 2018 Feb 20;18(1):126 [FREE Full text] [doi: 10.1186/s12913-018-2899-y]
46. Murray JL, Leung DT, Hanson OR, Ahmed SM, Pavia AT, Khan AI, et al. Drivers of inappropriate use of antimicrobials in South Asia: A systematic review of qualitative literature. PLOS Glob Public Health 2024 Apr 4;4(4):e0002507 [FREE Full text] [doi: 10.1371/journal.pgph.0002507]
47. Fishman N. Policy Statement on Antimicrobial Stewardship by the Society for Healthcare Epidemiology of America (SHEA), the Infectious Diseases Society of America (IDSA), and the Pediatric Infectious Diseases Society (PIDS). Infect. Control Hosp. Epidemiol 2015 Jan 02;33(4):322-327. [doi: 10.1086/665010]
48. Warreman EB, Lambregts MMC, Wouters RHP, Visser LG, Staats H, van Dijk E, et al. Determinants of in-hospital antibiotic prescription behaviour: a systematic review and formation of a comprehensive framework. Clinical Microbiology and Infection 2019 May;25(5):538-545. [doi: 10.1016/j.cmi.2018.09.006]
49. Step-by-step approach for development and implementation of hospital and antibiotic policy and standard treatment guidelines. World Health Organization. Regional Office for South-East Asia. 2011. URL: https://tinyurl.com/29pstpsh [accessed 2024-10-05]


## Abbreviations

**AMR:** antimicrobial resistance
**ASP:** antibiotic stewardship program
**CFI:** comparative fit index
**MTPB:** modified theory of planned behavior
**OPD:** outpatient department
**PBC:** perceived behavioral control
**SEM:** structural equation modeling
**SN:** subjective norm
**TPB:** theory of planned behavior
**URTI:** upper respiratory tract infection


*Edited by T de Azevedo Cardoso; submitted 11.02.24; peer-reviewed by T Shimelis, A Tsehay; comments to author 19.06.24; revised version received 22.06.24; accepted 24.09.24; published 23.10.24.*

*Please cite as:*
*Abejew AA, Wubetu GY, Fenta TG*
*Antibiotic Prescribing Behavior of Physicians in Outpatient Departments in Hospitals in Northwest Ethiopia: Structural Equation Modeling Approach*
*Interact J Med Res 2024;13:e57285*
*URL: https://www.i-jmr.org/2024/1/e57285*
*doi:10.2196/57285*
*PMID:39441643*

Original Paper

# Biochemical Changes in Adult Male Gamers During Prolonged Gaming: Pilot Study

Kasper Bygum Krarup[1], BSc, MD; Johannes Riis[1], MD; Morten Mørk[2], MD, PhD; Hien Thi Thu Nguyen[3], MSc, PhD; Inge Søkilde Pedersen[3,4], MSc, PhD; Søren Risom Kristensen[2,4], MD, PhD; Aase Handberg[2,4], MD, PhD; Henrik Bygum Krarup[3,4], MD, PhD

[1]Department of Geriatrics, Aalborg University Hospital, Aalborg, Denmark

[2]Department of Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark

[3]Department of Molecular Diagnostics, Aalborg University Hospital, Aalborg, Denmark

[4]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

**Corresponding Author:**
Kasper Bygum Krarup, BSc, MD
Department of Geriatrics
Aalborg University Hospital
Reberbansgade 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.

*(Interact J Med Res 2024;13:e46570)* doi:10.2196/46570

**KEYWORDS**

long gaming sessions; local area network party; biochemistry; cortisol; glucose; gaming; biochemical; blood sample; hematology; hematological; games; gamers; hemoglobin; adults; males; men; blood

## 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 adjourning 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/1875×*g* 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 (nmol/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. $sO_2$ (%), $pCO_2$ (kPa), $pO_2$ (kPa), pH, and standard bicarbonate concentration (mmol/L) were measured to assess acid-base balance and blood gases. Erythrocytes ($10^{12}$/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 ($10^{9}$/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 α (originally .05) by the number of tests. Statistical significance was set at α=.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.

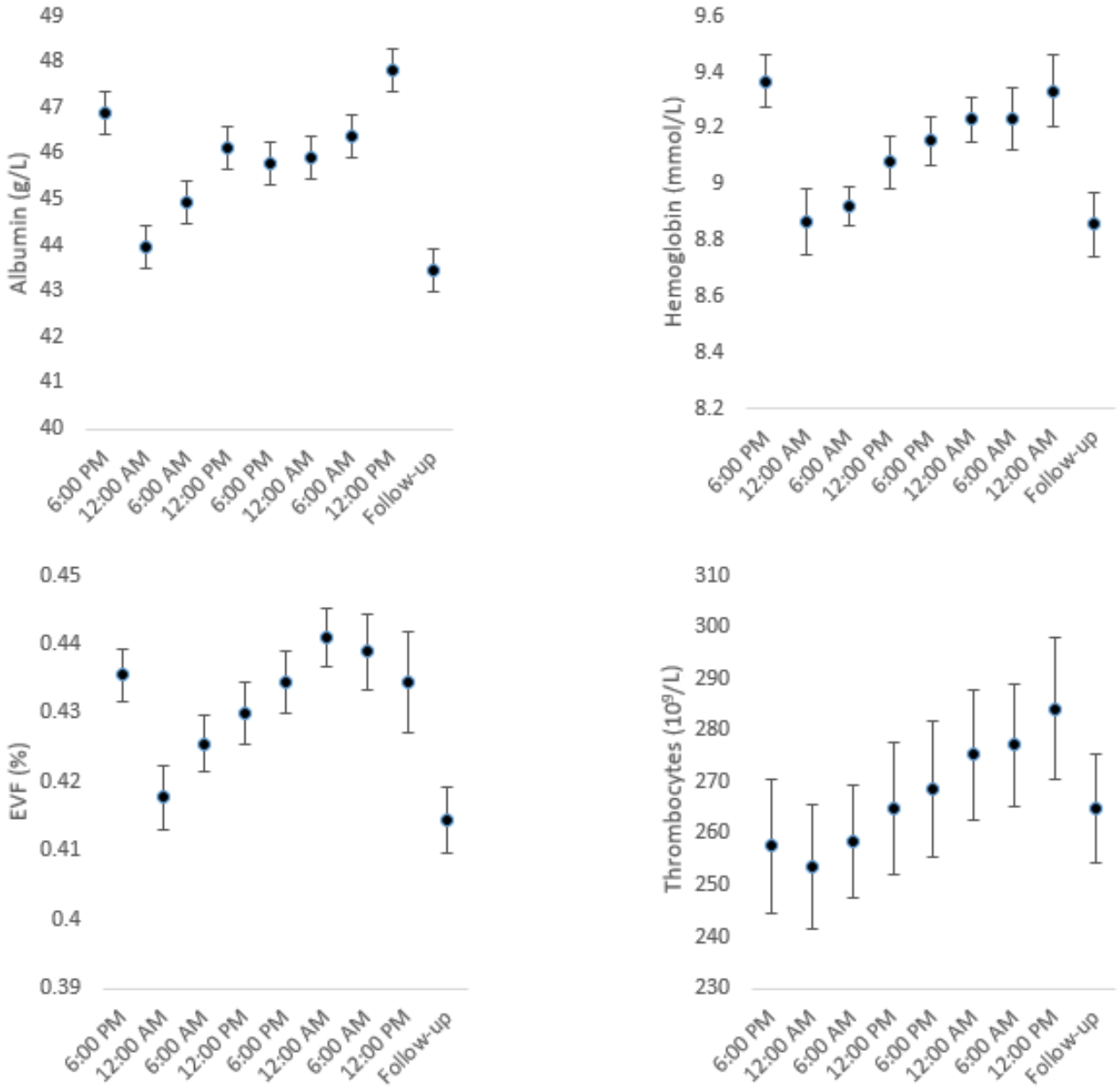


**Figure 2.** Glucose development over time.

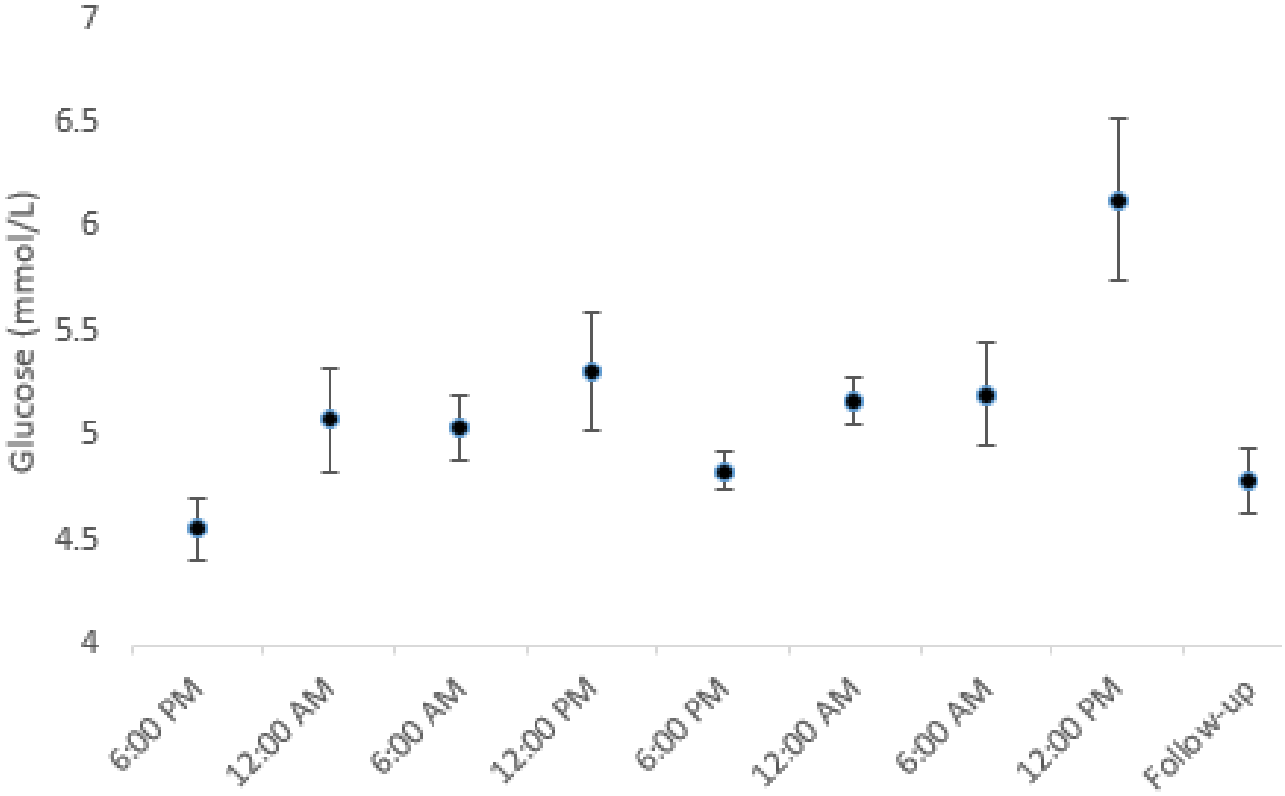



**Figure 3.** Cortisol development over time.

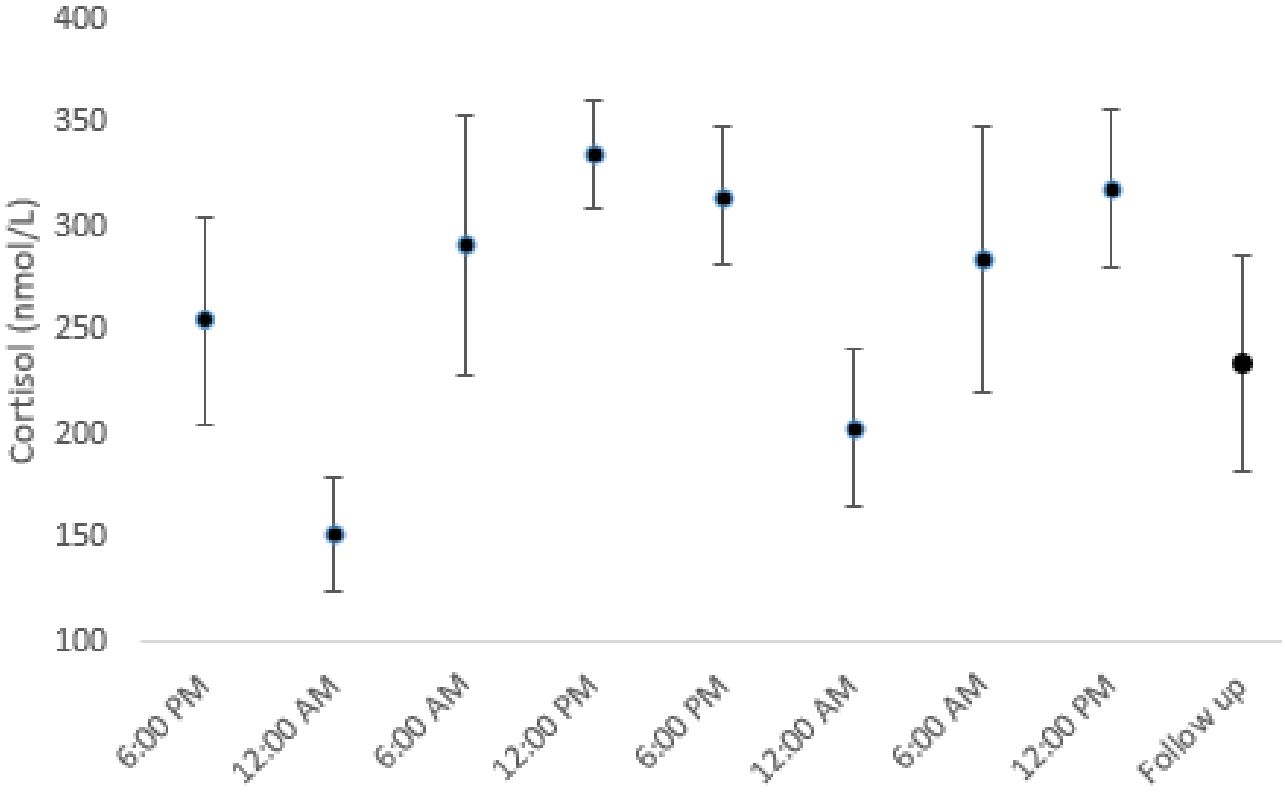
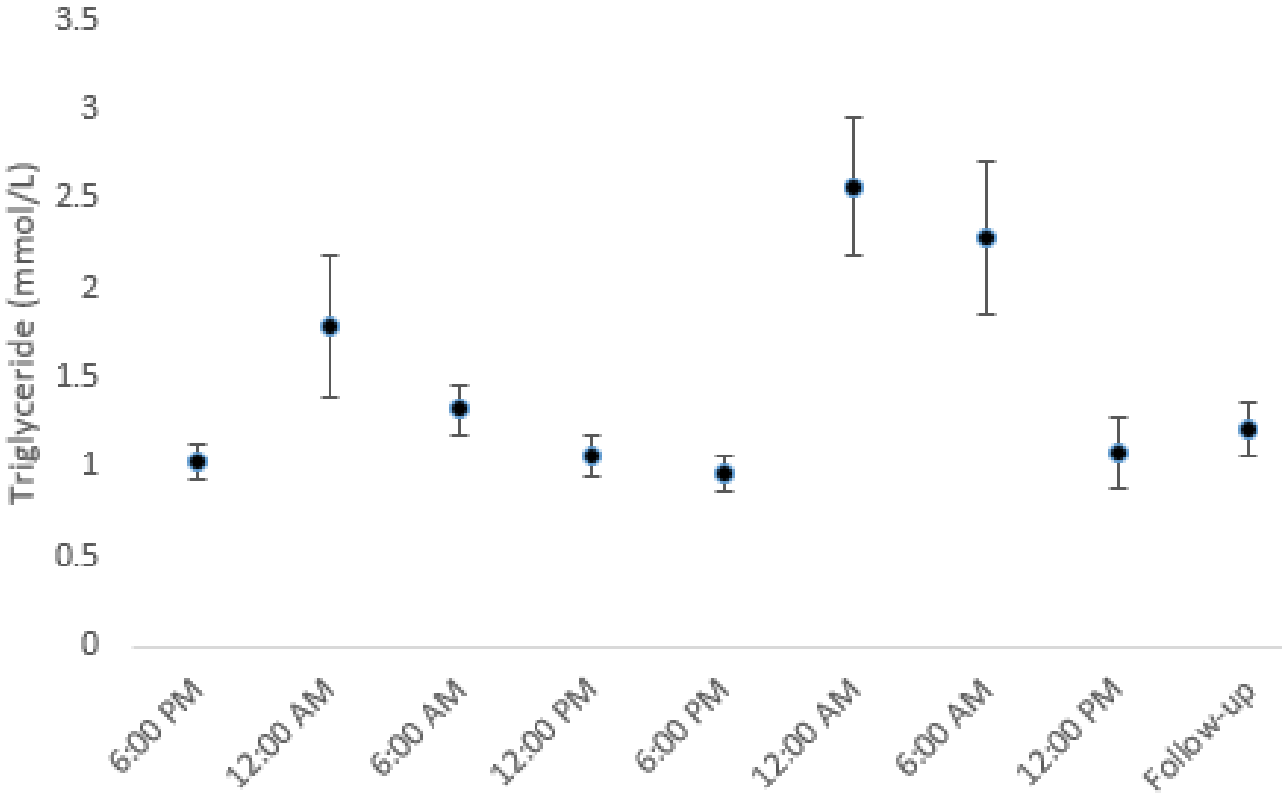


**Figure 4.** Triglyceride development over time.




## 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 every 6 hours during the first period and decreased by 1.15 (SD 0.56) µmol/L 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

$sO_2$, $pCO_2$, $pO_2$, 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 $258\times10^9$/L (SE $1.1\times10^9$/L) at baseline and increased by $6.9\times10^9$/L (SD $1.1\times10^9$/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 [13]. 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 $pO_2$ 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.

## Acknowledgments

The authors thank the participants for their input and participation in this study. This study would not have been possible without the support of Aalborg University Hospital, which generously provided facilities and technical support. Additionally, the authors thank the laboratory staff at the Department of Molecular Diagnostics and the Department of Clinical Biochemistry for their swift and tireless work during the study. This study was funded by grants from “Brødrene Hartmanns Fond” and “Tømmerhandler Johannes Fogs Fond.” The funding sources had no involvement in the study design, data collection, or data interpretation.

## Data Availability

Data from this study are not available in a public repository. It can be requested from the corresponding author. The study garnered national media attention in Denmark when it was conducted, and several participants were interviewed through radio and television. The data included physical characteristics, employment status, and other personal information. Due to the low number of participants, it would potentially be possible to identify individual participants’ data.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT checklist.
[PDF File (Adobe PDF File), 94 KB - ijmr_v13i1e46570_app1.pdf ]

Multimedia Appendix 2
Biochemical markers by analysis instrument.
[DOCX File , 16 KB - ijmr_v13i1e46570_app2.docx ]

Multimedia Appendix 3
Overview of biochemical markers.

[DOCX File , 55 KB - ijmr_v13i1e46570_app3.docx ]

## References


1. Lenhart J, Kahne J, Middaugh E, Macgill AR, Evans C, Vitak J. Teens, video games, and civics. Pew Research Center. 2008. URL: https://www.pewresearch.org/internet/2008/09/16/teens-video-games-and-civics/ [accessed 2024-04-19]
2. Turel O, Romashkin A, Morrison KM. Health outcomes of information system use lifestyles among adolescents: videogame addiction, sleep curtailment and cardio-metabolic deficiencies. PLoS One 2016;11(5):e0154764 [FREE Full text] [doi: 10.1371/journal.pone.0154764] [Medline: 27149512]
3. Turel O, Romashkin A, Morrison KM. A model linking video gaming, sleep quality, sweet drinks consumption and obesity among children and youth. Clin Obes 2017;7(4):191-198. [doi: 10.1111/cob.12191] [Medline: 28320073]
4. Vella CA, Taylor K, Nelson MC. Associations of leisure screen time with cardiometabolic biomarkers in college-aged adults. J Behav Med 2020;43(6):1014-1025 [FREE Full text] [doi: 10.1007/s10865-020-00161-2] [Medline: 32451650]
5. Vicente-Rodríguez G, Rey-López JP, Martín-Matillas M, Moreno LA, Wärnberg J, Redondo C, et al. Television watching, videogames, and excess of body fat in Spanish adolescents: the AVENA study. Nutrition 2008;24(7-8):654-662. [doi: 10.1016/j.nut.2008.03.011] [Medline: 18472399]
6. Vuong B, Rogers R, Corriveau N, Foti A, Pew A, Wei WC, et al. Passive screen time associated with unhealthy dietary consumption and physiological characteristics: a closer look at childhood behaviors. J Am Coll Cardiol 2014;63(12):A1292. [doi: 10.1016/s0735-1097(14)61292-4]
7. DiFrancisco-Donoghue J, Werner WG, Douris PC, Zwibel H. Esports players, got muscle? Competitive video game players' physical activity, body fat, bone mineral content, and muscle mass in comparison to matched controls. J Sport Health Sci 2022;11(6):725-730 [FREE Full text] [doi: 10.1016/j.jshs.2020.07.006] [Medline: 32711155]
8. Strasburger VC, Jordan AB, Donnerstein E. Children, adolescents, and the media: health effects. Pediatr Clin North Am 2012;59(3):533-587, vii. [doi: 10.1016/j.pcl.2012.03.025] [Medline: 22643165]
9. Henderson M, Benedetti A, Barnett TA, Mathieu ME, Deladoëy J, Gray-Donald K. Influence of adiposity, physical activity, fitness, and screen time on insulin dynamics over 2 years in children. JAMA Pediatr 2016;170(3):227-235 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.3909] [Medline: 26857733]
10. Siervo M, Gan J, Fewtrell MS, Cortina-Borja M, Wells JCK. Acute effects of video-game playing versus television viewing on stress markers and food intake in overweight and obese young men: a randomised controlled trial. Appetite 2018;120:100-108. [doi: 10.1016/j.appet.2017.08.018] [Medline: 28843974]
11. Mario S, Hannah C, Jonathan WCK, Jose L. Frequent video-game playing in young males is associated with central adiposity and high-sugar, low-fibre dietary consumption. Eat Weight Disord 2014;19(4):515-520. [doi: 10.1007/s40519-014-0128-1] [Medline: 24821488]
12. Gheller BJF, Totosy de Zepetnek JO, Welch JM, Rossiter MD, Luhovyy B, Brett NR, et al. Effect of video game playing and a glucose preload on subjective appetite, subjective emotions, and food intake in overweight and obese boys. Appl Physiol Nutr Metab 2019;44(3):248-254. [doi: 10.1139/apnm-2018-0281] [Medline: 30096245]
13. Krarup KB, Krarup HB. The physiological and biochemical effects of gaming: a review. Environ Res 2020;184:109344. [doi: 10.1016/j.envres.2020.109344] [Medline: 32199319]
14. Krarup KB, Rantanen JM, Mørk M, Pedersen IS, Christensen JH, Kristensen SR, et al. The physiological and cardiologic effects of long video gaming sessions in adult males. Sports Med Int Open 2022;6(1):E39-E46 [FREE Full text] [doi: 10.1055/a-1858-8436] [Medline: 35874049]
15. Chaput JP, Visby T, Nyby S, Klingenberg L, Gregersen NT, Tremblay A, et al. Video game playing increases food intake in adolescents: a randomized crossover study. Am J Clin Nutr 2011;93(6):1196-1203 [FREE Full text] [doi: 10.3945/ajcn.110.008680] [Medline: 21490141]
16. Wang X, Perry AC. Metabolic and physiologic responses to video game play in 7- to 10-year-old boys. Arch Pediatr Adolesc Med 2006;160(4):411-415 [FREE Full text] [doi: 10.1001/archpedi.160.4.411] [Medline: 16585487]
17. Oxford J, Ponzi D, Geary DC. Hormonal responses differ when playing violent video games against an ingroup and outgroup. Evol Hum Behav 2010;31(3):201-209. [doi: 10.1016/j.evolhumbehav.2009.07.002]
18. Kindermann H, Javor A, Reuter M. Playing counter-strike versus running: the impact of leisure time activities and cortisol on intermediate-term memory in male students. Cogn Syst Res 2016;40:1-7. [doi: 10.1016/j.cogsys.2016.01.002]
19. Ivarsson M, Anderson M, Akerstedt T, Lindblad F. Playing a violent television game does not affect saliva cortisol. Acta Paediatr 2009;98(6):1052-1053. [doi: 10.1111/j.1651-2227.2009.01271.x] [Medline: 19397536]
20. Dieckmann L. TV2 Nord, 2018. TV2 Nord. 2018 Sep 29. URL: https://vimeo.com/906653219 [accessed 2024-04-17]
21. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
22. Jafari M, Ansari-Pour N. Why, when and how to adjust your P values? Cell J 2019;20(4):604-607 [FREE Full text] [doi: 10.22074/cellj.2019.5992] [Medline: 30124010]
23. Giskeødegård GF, Lydersen S. Measurements below the detection limit. Tidsskr Nor Laegeforen 2022;142(13):62352 [FREE Full text] [doi: 10.4045/tidsskr.22.0439] [Medline: 36164797]

24. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
25. Seal AD, Bardis CN, Gavrieli A, Grigorakis P, Adams JD, Arnaoutis G, et al. Coffee with high but not low caffeine content augments fluid and electrolyte excretion at rest. Front Nutr 2017;4:40 [FREE Full text] [doi: 10.3389/fnut.2017.00040] [Medline: 28868290]
26. Röder PV, Wu B, Liu Y, Han W. Pancreatic regulation of glucose homeostasis. Exp Mol Med 2016;48(3):e219 [FREE Full text] [doi: 10.1038/emm.2016.6] [Medline: 26964835]
27. Ryan R, Booth S, Spathis A, Mollart S, Clow A. Use of salivary diurnal cortisol as an outcome measure in randomised controlled trials: a systematic review. Ann Behav Med 2016;50(2):210-236 [FREE Full text] [doi: 10.1007/s12160-015-9753-9] [Medline: 27007274]
28. Kudielka BM, Gierens A, Hellhammer DH, Wüst S, Schlotz W. Salivary cortisol in ambulatory assessment—some dos, some don'ts, and some open questions. Psychosom Med 2012;74(4):418-431. [doi: 10.1097/PSY.0b013e31825434c7] [Medline: 22582339]
29. Hellhammer DH, Wüst S, Kudielka BM. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinol 2009;34(2):163-171. [doi: 10.1016/j.psyneuen.2008.10.026] [Medline: 19095358]
30. Petrowski K, Buehrer S, Niedling M, Schmalbach B. The effects of light exposure on the cortisol stress response in human males. Stress 2021;24(1):29-35 [FREE Full text] [doi: 10.1080/10253890.2020.1741543] [Medline: 32160826]
31. Mirick DK, Bhatti P, Chen C, Nordt F, Stanczyk FZ, Davis S. Night shift work and levels of 6-sulfatoxymelatonin and cortisol in men. Cancer Epidemiol Biomarkers Prev 2013;22(6):1079-1087 [FREE Full text] [doi: 10.1158/1055-9965.EPI-12-1377] [Medline: 23563887]
32. Turel O. Videogames and guns in adolescents: tests of a bipartite theory. Comput Human Behav 2020;109:106355 [FREE Full text] [doi: 10.1016/j.chb.2020.106355] [Medline: 32372846]
33. Tiihonen K, Rautonen N, Alhoniemi E, Ahotupa M, Stowell J, Vasankari T. Postprandial triglyceride response in normolipidemic, hyperlipidemic and obese subjects—the influence of polydextrose, a non-digestible carbohydrate. Nutr J 2015;14:23 [FREE Full text] [doi: 10.1186/s12937-015-0009-0] [Medline: 25889643]
34. Mørk M, Nielsen MH, Bæk R, Jørgensen MM, Pedersen S, Kristensen SR. Postprandial increase in blood plasma levels of tissue factor-bearing (and other) microvesicles measured by flow cytometry: fact or artifact? TH Open 2018;2(2):e147-e157 [FREE Full text] [doi: 10.1055/s-0038-1642021] [Medline: 31249938]
35. Modi AA, Feld JJ, Park Y, Kleiner DE, Everhart JE, Liang TJ, et al. Increased caffeine consumption is associated with reduced hepatic fibrosis. Hepatology 2010;51(1):201-209 [FREE Full text] [doi: 10.1002/hep.23279] [Medline: 20034049]
36. Amer MG, Mazen NF, Mohamed AM. Caffeine intake decreases oxidative stress and inflammatory biomarkers in experimental liver diseases induced by thioacetamide: biochemical and histological study. Int J Immunopathol Pharmacol 2017;30(1):13-24 [FREE Full text] [doi: 10.1177/0394632017694898] [Medline: 28281876]
37. Helal MG, Ayoub SE, Elkashefand WF, Ibrahim TM. Caffeine affects HFD-induced hepatic steatosis by multifactorial intervention. Hum Exp Toxicol 2018;37(9):983-990 [FREE Full text] [doi: 10.1177/0960327117747026] [Medline: 29249184]
38. Marventano S, Salomone F, Godos J, Pluchinotta F, Del Rio D, Mistretta A, et al. Coffee and tea consumption in relation with non-alcoholic fatty liver and metabolic syndrome: a systematic review and meta-analysis of observational studies. Clin Nutr 2016;35(6):1269-1281. [doi: 10.1016/j.clnu.2016.03.012] [Medline: 27060021]
39. Molloy JW, Calcagno CJ, Williams CD, Jones FJ, Torres DM, Harrison SA. Association of coffee and caffeine consumption with fatty liver disease, nonalcoholic steatohepatitis, and degree of hepatic fibrosis. Hepatology 2012;55(2):429-436 [FREE Full text] [doi: 10.1002/hep.24731] [Medline: 21987293]
40. Kechagias S, Ernersson A, Dahlqvist O, Lundberg P, Lindström T, Nystrom FH, Fast Food Study Group. Fast-food-based hyper-alimentation can induce rapid and profound elevation of serum alanine aminotransferase in healthy subjects. Gut 2008;57(5):649-654 [FREE Full text] [doi: 10.1136/gut.2007.131797] [Medline: 18276725]


## 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

**TG:** triglyceride

*Edited by T de Azevedo Cardoso; submitted 16.02.23; peer-reviewed by X Yao, M Birk; comments to author 15.11.23; revised version received 30.01.24; accepted 02.04.24; published 08.07.24.*

*Please cite as:*
*Krarup KB, Riis J, Mørk M, Nguyen HTT, Søkilde Pedersen I, Risom Kristensen S, Handberg A, Krarup HB*
*Biochemical Changes in Adult Male Gamers During Prolonged Gaming: Pilot Study*
*Interact J Med Res 2024;13:e46570*
*URL: https://www.i-jmr.org/2024/1/e46570*
*doi:10.2196/46570*
*PMID:38976326*

Original Paper

# Effectiveness of a Smartphone App to Promote Physical Activity Among Persons With Type 2 Diabetes: Randomized Controlled Trial

Stephanie E Bonn[1], PhD; Madeleine Hummel[1], MD; Giulia Peveri[2,3], PhD; Helén Eke[1], MSc; Christina Alexandrou[4], PhD; Rino Bellocco[2,5], PhD; Marie Löf[4], PhD; Ylva Trolle Lagerros[1,6], MD, PhD

[1]Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[3]Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy

[4]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

[5]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy

[6]Center for Obesity, Academic Specialist Center, Stockholm Health Services, Stockholm, Sweden

**Corresponding Author:**
Stephanie E Bonn, PhD
Clinical Epidemiology Division, Department of Medicine Solna
Karolinska Institutet
Maria Aspmans gata 30A
Stockholm, 171 64
Sweden
Phone: 46 8 517 791
Email: stephanie.bonn@ki.se

## 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 $A_{1c}$, 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/$m^2$, 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

***(Interact J Med Res 2024;13:e53054)*** doi:10.2196/53054

**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 $A_{1c}$ ($HbA_{1c}$) 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, $HbA_{1c}$, 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, $HbA_{1c}$, 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 $HbA_{1c}$ 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 $HbA_{1c}$. The app was continuously 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, 4×15 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, $HbA_{1c}$ (mmol/mol), total cholesterol (mmol/L), LDL cholesterol (mmol/L), HDL cholesterol (mmol/L), and triglycerides (mmol/L) were measured in fasting blood samples. $HbA_{1c}$ 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 ($kg/m^2$).

## 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 continuously allocated to the next available spot on the list by study personnel (authors SEB and CA). Participants were informed about their group allocation at the end of the baseline meeting. Because of the nature of 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, $HbA_{1c}$, 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

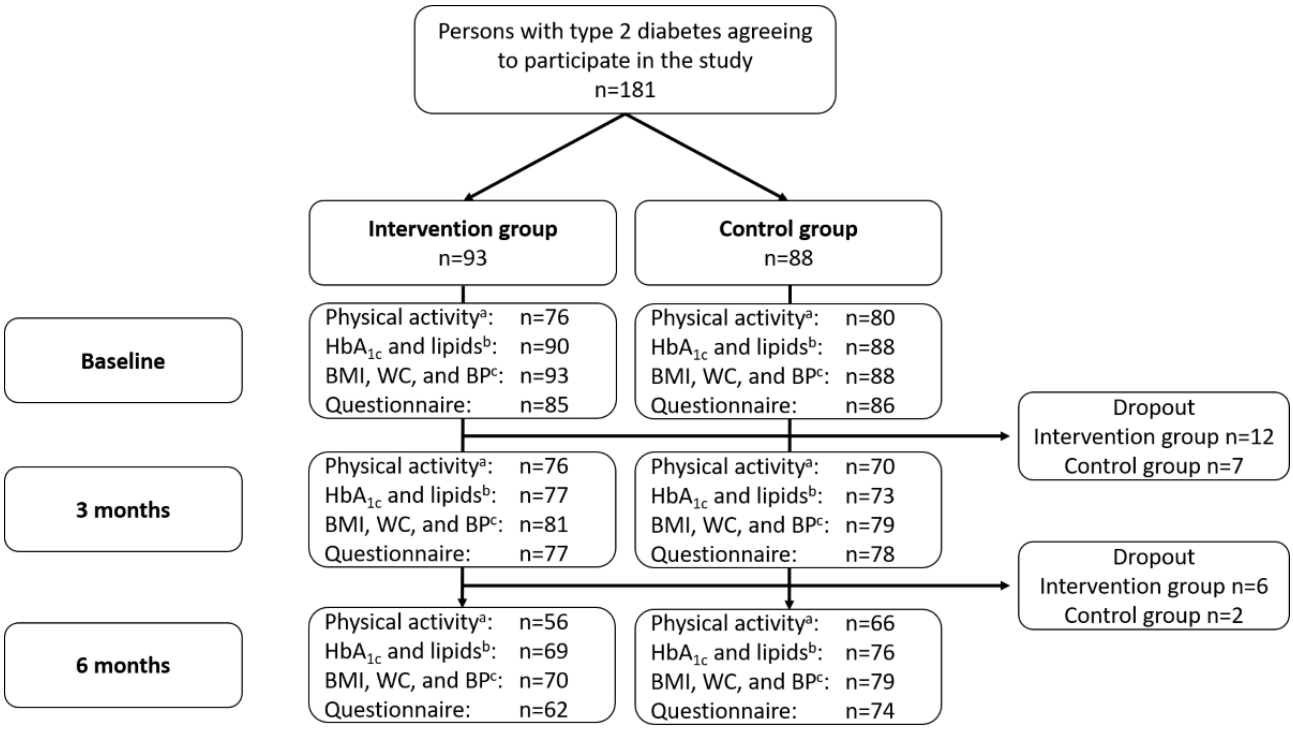
to contrast differences in MVPA between the intervention and control group using the imputed data for MVPA at baseline.

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. [a]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; [b]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 ≥25 kg/m$^2$, and 48.1% (87/181) had a BMI of ≥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).

| Characteristic | Intervention group (n=93) | | Control group (n=88) | | *P* value[a] |
|---|---|---|---|---|---|
| | n (%) | Mean (SD) | n (%) | Mean (SD) | |
| MVPA[b,c] (minutes/day) | 76 (82) | 38.3 (28.3) | 80 (91) | 29.7 (24.1) | *.04* |
| BMI (kg/m$^2$) | 93 (100) | 30.2 (5.5) | 88 (100) | 30.6 (5.2) | .61 |
| **Waist circumference** | | | | | |
| Women | 33 (35) | 102 (12.6) | 28 (32) | 103 (15.8) | .75 |
| Men | 59 (63) | 111 (15.6) | 60 (68) | 110 (13.3) | .75 |
| Hemoglobin $A_{1c}$ (mmol/mol) | 89 (96) | 53.6 (13.0) | 88 (100) | 53.5 (12.7) | .97 |
| Total cholesterol (mmol/L) | 74 (80) | 4.56 (1.00) | 70 (80) | 4.50 (1.18) | .77 |
| LDL[d] cholesterol (mmol/L) | 73 (78) | 2.70 (0.89) | 70 (80) | 2.64 (1.13) | .72 |
| HDL[e] cholesterol (mmol/L) | 74 (80) | 1.25 (0.34) | 70 (80) | 1.23 (0.40) | .73 |
| Triglycerides (mmol/L) | 72 (77) | 1.54 (0.76) | 69 (78) | 1.64 (0.94) | .45 |
| **Blood pressure (mm Hg)** | | | | | |
| Systolic | 92 (99) | 138.7 (16.2) | 88 (100) | 137.0 (14.8) | .45 |
| Diastolic | 92 (99) | 83.6 (9.9) | 88 (100) | 82.6 (9.5) | .50 |
| **Sex** | | | | | .50 |
| Male | 34 (37) | —[f] | 28 (32) | — | |
| Female | 59 (63) | — | 60 (68) | — | |
| **Age (years)** | | | | | .24 |
| <50 | 20 (22) | — | 16 (18) | — | |
| 50-59 | 28 (30) | — | 19 (22) | — | |
| 60-69 | 31 (33) | — | 30 (34) | — | |
| ≥70 | 14 (15) | — | 23 (26) | — | |
| **Leisure time activity[g] (minutes/week)** | | | | | .27 |
| <60 | 6 (7) | | 11 (13) | — | |
| 60-90 | 8 (10) | — | 8 (10) | — | |
| 90-150 | 17 (20) | — | 15 (18) | — | |
| 150-300 | 14 (17) | — | 22 (26) | — | |
| >300 | 39 (46) | — | 28 (33) | — | |
| **Primary care centers** | | | | | .88 |
| 1 | 35 (38) | — | 31 (35) | — | |
| 2 | 14 (15) | — | 13 (15) | — | |
| 3 | 6 (6) | — | 8 (9) | — | |
| 4 | 23 (25) | — | 23 (26) | — | |
| 5 | 10 (11) | — | 11 (13) | — | |
| Specialized medical center | 5 (5) | — | 2 (2) | — | |
| **Time spent exercising[h] (minutes/week)** | | | | | .26 |
| Never | 35 (42) | — | 38 (44) | — | |
| <30 | 12 (14) | — | 12 (14) | — | |
| 30-90 | 13 (15) | — | 21 (24) | — | |

| Characteristic | Intervention group (n=93) | | Control group (n=88) | | *P* value[a] |
|---|---|---|---|---|---|
| | n (%) | Mean (SD) | n (%) | Mean (SD) | |
| >90 | 24 (29) | — | 15 (17) | — | |
| **Smoking[i]** | | | | | .41 |
| Yes | 11 (13) | — | 9 (11) | — | |
| No, ever smoker | 31 (37) | — | 40 (47) | — | |
| No, never smoker | 42 (50) | — | 36 (42) | — | |
| **Time since diabetes diagnosis[j] (years)** | | | | | .50 |
| <1 | 8 (11) | — | 13 (18) | — | |
| 1-5 | 20 (29) | — | 17 (24) | — | |
| >5 | 42 (60) | — | 42 (58) | — | |
| **Education[k] (years)** | | | | | .69 |
| ≤12 | 45 (54) | — | 44 (51) | — | |
| >12 | 38 (46) | — | 42 (49) | — | |

[a]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.

[b]MVPA: moderate to vigorous physical activity.

[c]Missing data from n=17 (intervention) and n=8 (control).

[d]LDL: low-density lipoprotein.

[e]HDL: high-density lipoprotein.

[f]—not available.

[g]From questionnaire, missing data n=9 (intervention) and n=4 (control).

[h]From questionnaire, missing data n=9 (intervention) and n=2 (control).

[i]Missing data n=9 (intervention) and n=3 (control).

[j]Missing data n=23 (intervention) and n=16 (control).

[k]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.

| Characteristic | Group sample means | | | | Model estimates[a] | |
|---|---|---|---|---|---|---|
| | Intervention (n=93) | | Control (n=88) | | Difference[b] | Group by time interaction |
| | n (%) | Mean (SD) | n (%) | Mean (SD) | Mean (95% CI) | β (95% CI) |
| **MVPA (minutes/day)** | | | | | | |
| 3 months | 70 (75) | 36.6 (25.5) | 67 (76) | 26.7 (21.1) | 10.05 (1.66 to 18.44) | 1.51 (–5.53 to 8.55) |
| 6 months | 55 (59) | 34.2 (29.4) | 63 (72) | 31.1 (27.0) | 5.02 (–3.72 to 13.75) | –3.53 (–10.97 to 3.92) |
| **BMI (kg/m$^2$)** | | | | | | |
| Baseline | 93 (100) | 30.2 (5.5) | 88 (100) | 30.6 (5.2) | N/A[c] | N/A |
| 3 months | 81 (87) | 30.1 (5.7) | 79 (90) | 30.3 (5.0) | –0.12 (–1.67 to 1.44) | 0.29 (–0.02 to 0.61) |
| 6 months | 70 (75) | 30.0 (6.0) | 79 (90) | 29.9 (5.0) | 0.11 (–1.45 to 1.67) | 0.52 (0.20 to 0.84) |
| **Waist circumference (cm)** | | | | | | |
| Baseline | 92 (99) | 107 (15.1) | 88 (100) | 108 (14.4) | N/A | N/A |
| 3 months | 80 (86) | 107 (15.6) | 79 (90) | 108 (13.6) | –0.61 (–4.90 to 3.69) | –0.46 (–1.75 to 0.83) |
| 6 months | 69 (74) | 107 (17.0) | 79 (90) | 106 (13.6) | 0.47 (–3.84 to 4.77) | 0.61 (–0.71 to 1.94) |
| **Hemoglobin $A_{1c}$ (mmol/mol)** | | | | | | |
| Baseline | 89 (96) | 53.6 (13.0) | 88 (100) | 53.5 (12.7) | N/A | N/A |
| 3 months | 75 (81) | 50.0 (9.9) | 73 (83) | 53.2 (13.4) | –2.45 (–6.08 to 1.17) | –2.54 (–5.36 to 0.29) |
| 6 months | 67 (72) | 51.2 (10.7) | 76 (86) | 51.2 (10.6) | –0.21 (–3.87 to 3.44) | –0.30 (–3.16 to 2.57) |
| **Total cholesterol (mmol/L)** | | | | | | |
| Baseline | 74 (80) | 4.56 (1.00) | 70 (80) | 4.50 (1.18) | N/A | N/A |
| 3 months | 62 (67) | 4.39 (0.80) | 57 (65) | 4.48 (0.91) | –0.06 (–0.40 to 0.28) | –0.11 (–0.37 to 0.14) |
| 6 months | 55 (59) | 4.38 (0.84) | 59 (67) | 4.27 (1.02) | 0.10 (–0.24 to 0.44) | 0.05 (–0.21 to 0.31) |
| **LDL[d] cholesterol (mmol/L)** | | | | | | |
| Baseline | 73 (78) | 2.70 (0.89) | 70 (80) | 2.64 (1.13) | N/A | N/A |
| 3 months | 59 (63) | 2.44 (0.65) | 56 (64) | 2.48 (0.89) | –0.01 (–0.33 to 0.30) | –0.07 (–0.30 to 0.15) |
| 6 months | 54 (58) | 2.44 (0.68) | 57 (65) | 2.38 (0.85) | 0.09 (–0.23 to 0.40) | 0.02 (–0.20 to 0.25) |
| **HDL[e] cholesterol (mmol/L)** | | | | | | |
| Baseline | 74 (80) | 1.25 (0.34) | 70 (80) | 1.23 (0.40) | N/A | N/A |
| 3 months | 62 (67) | 1.22 (0.34) | 57 (65) | 1.24 (0.38) | –0.04 (–0.16 to 0.09) | –0.06 (–0.12 to 0.01) |
| 6 months | 55 (59) | 1.26 (0.36) | 57 (65) | 1.24 (0.39) | –0.04 (–0.17 to 0.08) | –0.06 (–0.13 to 0.001) |
| **Triglycerides (mmol/L)** | | | | | | |
| Baseline | 72 (77) | 1.54 (0.76) | 69 (78) | 1.65 (0.94) | N/A | N/A |
| 3 months | 60 (65) | 1.71 (1.03) | 58 (66) | 1.76 (0.96) | –0.01 (–0.32 to 0.29) | 0.10 (–0.15 to 0.34) |
| 6 months | 52 (56) | 1.56 (0.83) | 56 (64) | 1.66 (0.83) | –0.07 (–0.38 to 0.24) | 0.04 (–0.21 to 0.29) |
| **Systolic BP[f] (mm Hg)** | | | | | | |
| Baseline | 92 (99) | 139 (16.2) | 88 (100) | 137 (14.8) | N/A | N/A |
| 3 months | 80 (86) | 135 (13.3) | 79 (90) | 134 (11.8) | 1.21 (–2.98 to 5.40) | –0.54 (–4.62 to 3.53) |
| 6 months | 70 (75) | 136 (12.8) | 78 (89) | 136 (12.6) | 0.70 (–3.60 to 4.99) | –1.05 (–5.23 to 3.13) |
| **Diastolic BP (mm Hg)** | | | | | | |
| Baseline | 92 (99) | 83.6 (9.9) | 88 (100) | 82.6 (9. 5) | N/A | N/A |
| 3 months | 80 (86) | 81.8 (9.1) | 79 (90) | 79.2 (10.5) | 2.78 (–0.14 to 5.69) | 1.79 (–1.06 to 4.64) |

| Characteristic | Group sample means | | | | Model estimates[a] | |
|---|---|---|---|---|---|---|
| | Intervention (n=93) | | Control (n=88) | | Difference[b] | Group by time interaction |
| | n (%) | Mean (SD) | n (%) | Mean (SD) | Mean (95% CI) | β (95% CI) |
| 6 months | 70 (75) | 81.2 (8.7) | 78 (89) | 79.9 (9.4) | 1.87 (–1.12 to 4.85) | 0.88 (–2.04 to 3.81) |

[a]Results from linear mixed model analysis including participants with complete baseline data.

[b]Difference between groups at the specified time point based on predicted means from linear mixed models.

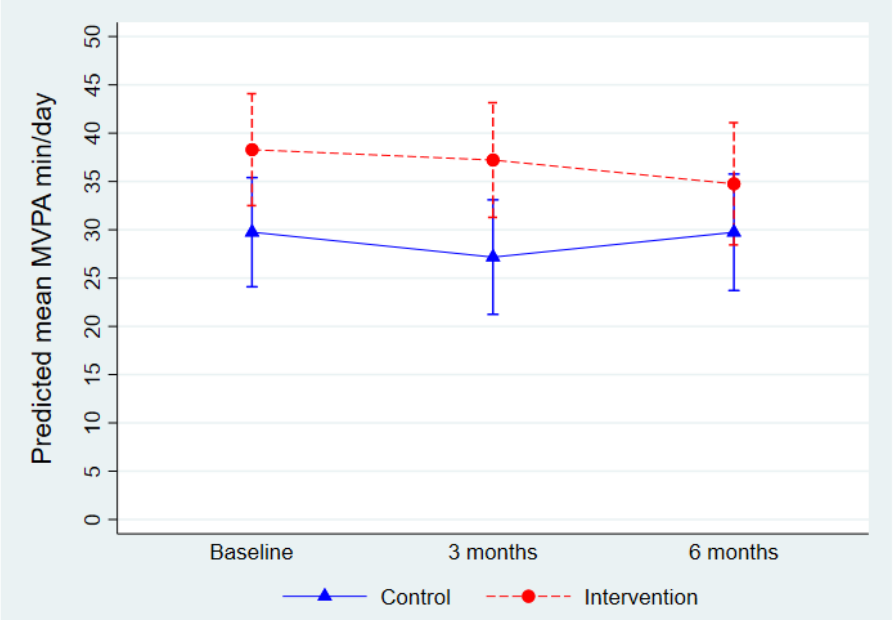
[c]—: not applicable.

[d]LDL: low-density lipoprotein.

[e]HDL: high-density lipoprotein.

[f]BP: blood pressure.

**Figure 2.** Changes over time in moderate to vigorous physical activity (MVPA, minutes/day) in the intervention and control group. Predicted group means from linear mixed model analysis. Results correspond to model estimates shown in Table 2.




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

the studies, as a higher baseline level of MVPA may imply less room for improvement.

Other previous studies have also shown an effect of mHealth solutions to increase physical activity. Hochsmann et al [35] showed that an interactive smartphone game aimed at increasing daily steps in persons with type 2 diabetes had a significant effect on activity compared with a control group receiving standard lifestyle counseling. Glynn et al [36] evaluated the effect of an app that aimed to increase physical activity in primary care patients and found that the intervention group increased their number of daily steps compared with the control group. In the randomized Sophia step study, comprising persons with prediabetes or type 2 diabetes in Sweden [37], self-monitoring of daily physical activity using pedometers with registration of steps on the web in combination with counseling did not increase levels of physical activity but seemed to prevent the decrease in physical activity seen in the control group. Although our primary hypothesis was rejected and we could not show that the use of the DiaCert app led to increased MVPA among persons with type 2 diabetes, results from other studies still indicate that mHealth solutions can have positive effects on physical activity.

Although we found a statistically significant effect of the DiaCert intervention on BMI at the 6-month follow-up, there was no difference in mean BMI between the groups at any time point. Therefore, we cannot draw any conclusion regarding the effect on BMI. Not all studies evaluating interventions targeting physical activity include anthropometric outcomes, and the effect on BMI in previous studies are mixed, showing no effect or indicating a small reduction favoring the intervention group [10,12]. Thorsen et al [20] did not report BMI as an outcome, but found a nonsignificant reduction of waist circumference. Similar to the Sophia step study, our results provided no evidence for an effect on waist circumference [37].

Reviews and meta-analyses [11,13,15] provide strong evidence for a positive effect of mobile apps for lifestyle modification in persons with type 2 diabetes on $HbA_{1c}$ levels. However, all of the evaluated apps included monitoring of blood glucose. The effects on other clinical markers including body weight, blood lipids, and blood pressure are less clear [38]. Our results are in line with the Sophia step study, where no effects on cardiometabolic variables, including $HbA_{1c}$, triglycerides, HDL cholesterol, and LDL cholesterol, were seen [37]. Neither our study nor the Sophia step study included a glucose-monitoring component, which may explain the lack of an effect on $HbA_{1c}$. Nevertheless, results from a meta-analysis by Lee et al [12] indicated that mHealth interventions in persons with type 2 diabetes >65 years of age may improve blood lipid profiles, which we found no evidence of. One explanation for the lack of an effect on cardiometabolic markers is that a changed lifestyle behavior for the better could have led to lowered medication, leaving $HbA_{1c}$, serum lipid levels, or blood pressure unchanged. We did not assess changes in medications during the intervention, which is a limitation of our study.

The use of different behavior change techniques may also explain differences in efficacy between apps. Fanning et al [39] 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 positive 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 $HbA_{1c}$ 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 $HbA_{1c}$ 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, $HbA_{1c}$, 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

1. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes—global burden of disease and forecasted trends. J Epidemiol Glob Health 2020;10(1):107-111 [FREE Full text] [doi: 10.2991/jegh.k.191028.001] [Medline: 32175717]

2. American Diabetes Association. 5. Lifestyle management: standards of medical care in diabetes-2019. Diabetes Care 2019;42(Suppl 1):S46-S60 [FREE Full text] [doi: 10.2337/dc19-S005] [Medline: 30559231]
3. Kanaley JA, Colberg SR, Corcoran MH, Malin SK, Rodriguez NR, Crespo CJ, et al. Exercise/physical activity in individuals with type 2 diabetes: a consensus statement from the American College of Sports Medicine. Med Sci Sports Exerc 2022;54(2):353-368 [FREE Full text] [doi: 10.1249/MSS.0000000000002800] [Medline: 35029593]
4. Ried-Larsen M, MacDonald CS, Johansen MY, Hansen KB, Christensen R, Almdal TP, et al. Why prescribe exercise as therapy in type 2 diabetes? We have a pill for that!. Diabetes Metab Res Rev 2018;34(5):e2999. [doi: 10.1002/dmrr.2999] [Medline: 29488311]
5. Moghetti P, Balducci S, Guidetti L, Mazzuca P, Rossi E, Schena F, Italian Society of Diabetology (SID), Italian Association of Medical Diabetologists (AMD), Italian Society of Motor and Sports Sciences (SISMES). Walking for subjects with type 2 diabetes: a systematic review and joint AMD/SID/SISMES evidence-based practical guideline. Nutr Metab Cardiovasc Dis 2020;30(11):1882-1898 [FREE Full text] [doi: 10.1016/j.numecd.2020.08.021] [Medline: 32998820]
6. Karduck J, Chapman-Novakofski K. Results of the clinician apps survey, how clinicians working with patients with diabetes and obesity use mobile health apps. J Nutr Educ Behav 2018;50(1):62-69.e1. [doi: 10.1016/j.jneb.2017.06.004] [Medline: 29325664]
7. Jimenez G, Lum E, Car J. Examining diabetes management apps recommended from a Google search: content analysis. JMIR Mhealth Uhealth 2019;7(1):e11848 [FREE Full text] [doi: 10.2196/11848] [Medline: 30303485]
8. Sneha S, Thalla S, Rischie I, Shahriar H. Health internet technology for chronic conditions: review of diabetes management apps. JMIR Diabetes 2021;6(3):e17431 [FREE Full text] [doi: 10.2196/17431] [Medline: 34463627]
9. Bonn SE, Alexandrou C, Lagerros YT. A digital platform and smartphone app to increase physical activity in patients with type 2 diabetes: overview of a technical solution. JMIR Form Res 2023;7:e40285 [FREE Full text] [doi: 10.2196/40285] [Medline: 36917156]
10. Eberle C, Löhnert M, Stichling S. Effectiveness of disease-specific mHealth apps in patients with diabetes mellitus: scoping review. JMIR Mhealth Uhealth 2021;9(2):e23477 [FREE Full text] [doi: 10.2196/23477] [Medline: 33587045]
11. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
12. Lee JJN, Abdul Aziz A, Chan ST, Sahrizan RSFBRA, Ooi AYY, Teh YT, et al. Effects of mobile health interventions on health-related outcomes in older adults with type 2 diabetes: a systematic review and meta-analysis. J Diabetes 2023;15(1):47-57 [FREE Full text] [doi: 10.1111/1753-0407.13346] [Medline: 36649940]
13. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
14. Moschonis G, Siopis G, Jung J, Eweka E, Willems R, Kwasnicka D, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023;5(3):e125-e143 [FREE Full text] [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]
15. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
16. Feter N, Dos Santos TS, Caputo EL, da Silva MC. What is the role of smartphones on physical activity promotion? A systematic review and meta-analysis. Int J Public Health 2019;64(5):679-690. [doi: 10.1007/s00038-019-01210-7] [Medline: 30758514]
17. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
18. Pradal-Cano L, Lozano-Ruiz C, Pereyra-Rodríguez JJ, Saigí-Rubió F, Bach-Faig A, Esquius L, et al. Using mobile applications to increase physical activity: a systematic review. Int J Environ Res Public Health 2020;17(21):8238 [FREE Full text] [doi: 10.3390/ijerph17218238] [Medline: 33171871]
19. Poppe L, De Bourdeaudhuij I, Verloigne M, Shadid S, Van Cauwenberg J, Compernolle S, et al. Efficacy of a self-regulation-based electronic and mobile health intervention targeting an active lifestyle in adults having type 2 diabetes and in adults aged 50 years or older: two randomized controlled trials. J Med Internet Res 2019;21(8):e13363 [FREE Full text] [doi: 10.2196/13363] [Medline: 31376274]
20. Thorsen IK, Yang Y, Valentiner LS, Glümer C, Karstoft K, Brønd JC, et al. The effects of a lifestyle intervention supported by the InterWalk Smartphone App on increasing physical activity among persons with type 2 diabetes: parallel-group, randomized trial. JMIR Mhealth Uhealth 2022;10(9):e30602 [FREE Full text] [doi: 10.2196/30602] [Medline: 36170002]
21. Bonn SE, Alexandrou C, Steiner KH, Wiklander K, Östenson CG, Löf M, et al. App-technology to increase physical activity among patients with diabetes type 2—the DiaCert-study, a randomized controlled trial. BMC Public Health 2018;18(1):119 [FREE Full text] [doi: 10.1186/s12889-018-5026-4] [Medline: 29316905]

22. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152(11):726-732 [FREE Full text] [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
23. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
24. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
25. Migueles JH, Cadenas-Sanchez C, Ekelund U, Nyström CD, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017;47(9):1821-1845 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
26. Migueles JH, Huber F, Sabia S, Rowlands AV, van Hees VT. GGIR: a research community-driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. J Meas Phys Behav 2019;2(3):188-186 [FREE Full text] [doi: 10.1123/jmpb.2018-0063]
27. Menai M, van Hees VT, Elbaz A, Kivimaki M, Singh-Manoux A, Sabia S. Accelerometer assessed moderate-to-vigorous physical activity and successful ageing: results from the Whitehall II study. Sci Rep 2017;8:45772 [FREE Full text] [doi: 10.1038/srep45772] [Medline: 28367987]
28. Olsson SJG, Ekblom Ö, Andersson E, Börjesson M, Kallings LV. Categorical answer modes provide superior validity to open answers when asking for level of physical activity: a cross-sectional study. Scand J Public Health 2016;44(1):70-76. [doi: 10.1177/1403494815602830] [Medline: 26392418]
29. Jeppsson JO, Kobold U, Barr J, Finke A, Hoelzel W, Hoshino T, et al. Approved IFCC reference method for the measurement of HbA1c in human blood. Clin Chem Lab Med 2002;40(1):78-89. [doi: 10.1515/CCLM.2002.016] [Medline: 11916276]
30. Nilsson-Ehle P, Berggren So¨derlund M, Theodorsson E, Becker C, Laurell CB. Laurells Klinisk kemi i praktisk medicin. In: Laurells Klinisk kemi i praktisk medicin, 9. uppl [Laurell's Clinical Chemistry in Practical Medicine, 9th Edition]. Lund, Sweden: Studentlitteratur; 2012.
31. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972;18(6):499-502. [Medline: 4337382]
32. Twisk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M. Different ways to estimate treatment effects in randomised controlled trials. Contemp Clin Trials Commun 2018;10:80-85 [FREE Full text] [doi: 10.1016/j.conctc.2018.03.008] [Medline: 29696162]
33. Li P, Stuart EA. Best (but oft-forgotten) practices: missing data methods in randomized controlled nutrition trials. Am J Clin Nutr 2019;109(3):504-508 [FREE Full text] [doi: 10.1093/ajcn/nqy271] [Medline: 30793174]
34. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res 2011;20(1):40-49 [FREE Full text] [doi: 10.1002/mpr.329] [Medline: 21499542]
35. Höchsmann C, Müller O, Ambühl M, Klenk C, Königstein K, Infanger D, et al. Novel smartphone game improves physical activity behavior in type 2 diabetes. Am J Prev Med 2019;57(1):41-50 [FREE Full text] [doi: 10.1016/j.amepre.2019.02.017] [Medline: 31128953]
36. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014;64(624):e384-e391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
37. Rossen J, Larsson K, Hagströmer M, Yngve A, Brismar K, Ainsworth B, et al. Effects of a three-armed randomised controlled trial using self-monitoring of daily steps with and without counselling in prediabetes and type 2 diabetes-the Sophia Step Study. Int J Behav Nutr Phys Act 2021;18(1):121 [FREE Full text] [doi: 10.1186/s12966-021-01193-w] [Medline: 34496859]
38. Kim JE, Park TS, Kim KJ. The clinical effects of type 2 diabetes patient management using digital healthcare technology: a systematic review and meta-analysis. Healthcare (Basel) 2022;10(3):522 [FREE Full text] [doi: 10.3390/healthcare10030522] [Medline: 35327000]
39. Fanning J, Roberts S, Hillman CH, Mullen SP, Ritterband L, McAuley E. A smartphone "app"-delivered randomized factorial trial targeting physical activity in adults. J Behav Med 2017;40(5):712-729. [doi: 10.1007/s10865-017-9838-y] [Medline: 28255750]
40. Sporrel K, Nibbeling N, Wang S, Ettema D, Simons M. Unraveling mobile health exercise interventions for adults: scoping review on the implementations and designs of persuasive strategies. JMIR Mhealth Uhealth 2021;9(1):e16282 [FREE Full text] [doi: 10.2196/16282] [Medline: 33459598]
41. Nationella Diabetesregistret—ÅRsrapport 2020 [The Swedish National Diabetes Register, annual report 2020]. Nationella Diabetesregistret. 2020. URL: https://www.ndr.nu/pdfs/Arsrapport_NDR_2020.pdf [accessed 2024-02-13]
42. Annual report—The Swedes and the Internet 2021. Swedish Svenskarna och internet [The Swedish Internet Foundation]. 2021. URL: https://svenskarnaochinternet.se/english/earlier-years-reports/ [accessed 2024-02-13]
43. Stevens W, Hillsdon M, Thorogood M, McArdle D. Cost-effectiveness of a primary care based physical activity intervention in 45-74 year old men and women: a randomised controlled trial. Br J Sports Med 1998;32(3):236-241 [FREE Full text] [doi: 10.1136/bjsm.32.3.236] [Medline: 9773174]

44. Lavikainen P, Mattila E, Absetz P, Harjumaa M, Lindström J, Järvelä-Reijonen E, et al. Digitally supported lifestyle intervention to prevent type 2 diabetes through healthy habits: secondary analysis of long-term user engagement trajectories in a randomized controlled trial. J Med Internet Res 2022;24(2):e31530 [FREE Full text] [doi: 10.2196/31530] [Medline: 35200147]
45. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]

## Abbreviations

**CONSORT:** Consolidated Standards of Reporting Trials
**CONSORT-EHEALTH:** Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth
**$HbA_{1c}$:** hemoglobin $A_{1c}$
**HDL:** high-density lipoprotein
**LDL:** low-density lipoprotein
**mHealth:** mobile health
**MICE:** multiple imputation by chained equations
**MVPA:** moderate to vigorous physical activity

*Edited by T de Azevedo Cardoso; submitted 26.09.23; peer-reviewed by I Thorsen, A Mateo-Orcajada, G Wren; comments to author 15.11.23; revised version received 22.12.23; accepted 07.02.24; published 21.03.24.*

*Please cite as:*
*Bonn SE, Hummel M, Peveri G, Eke H, Alexandrou C, Bellocco R, Löf M, Trolle Lagerros Y*
*Effectiveness of a Smartphone App to Promote Physical Activity Among Persons With Type 2 Diabetes: Randomized Controlled Trial*
*Interact J Med Res 2024;13:e53054*
*URL: https://www.i-jmr.org/2024/1/e53054*
*doi:10.2196/53054*
*PMID:38512333*

Original Paper

# Consequences of Data Loss on Clinical Decision-Making in Continuous Glucose Monitoring: Retrospective Cohort Study

Niala den Braber[1,2*], MSc; Carlijn I R Braem[1,2*], MSc; Miriam M R Vollenbroek-Hutten[1], Prof Dr; Hermie J Hermens[1], Prof Dr Ir; Thomas Urgert[1,2], MSc; Utku S Yavuz[1], PhD; Peter H Veltink[1], Prof Dr Ir; Gozewijn D Laverman[1,2], Prof Dr, MD

[1]Biomedical Signal and Systems, Faculty of Electrical Engineering, Mathematics And Computer Science, University of Twente, Enschede, Netherlands
[2]Internal Medicine, Ziekenhuisgroep Twente, Almelo, Netherlands
[*]these authors contributed equally

**Corresponding Author:**
Carlijn I R Braem, MSc
Biomedical Signal and Systems
Faculty of Electrical Engineering, Mathematics And Computer Science
University of Twente
PO BOX 217
Enschede, 7500 AE
Netherlands
Phone: 31 534898830
Email: c.i.r.braem@utwente.nl

## Abstract

**Background:** The impact of missing data on individual continuous glucose monitoring (CGM) data is unknown but can influence clinical decision-making for patients.

**Objective:** We aimed to investigate the consequences of data loss on glucose metrics in individual patient recordings from continuous glucose monitors and assess its implications on clinical decision-making.

**Methods:** The CGM data were collected from patients with type 1 and 2 diabetes using the FreeStyle Libre sensor (Abbott Diabetes Care). We selected 7-28 days of 24 hours of continuous data without any missing values from each individual patient. To mimic real-world data loss, missing data ranging from 5% to 50% were introduced into the data set. From this modified data set, clinical metrics including time below range (TBR), TBR level 2 (TBR2), and other common glucose metrics were calculated in the data sets with and that without data loss. Recordings in which glucose metrics deviated relevantly due to data loss, as determined by clinical experts, were defined as expert panel boundary error ($\varepsilon_{EPB}$). These errors were expressed as a percentage of the total number of recordings. The errors for the recordings with glucose management indicator <53 mmol/mol were investigated.

**Results:** A total of 84 patients contributed to 798 recordings over 28 days. With 5%-50% data loss for 7-28 days recordings, the $\varepsilon_{EPB}$ varied from 0 out of 798 (0.0%) to 147 out of 736 (20.0%) for TBR and 0 out of 612 (0.0%) to 22 out of 408 (5.4%) recordings for TBR2. In the case of 14-day recordings, TBR and TBR2 episodes completely disappeared due to 30% data loss in 2 out of 786 (0.3%) and 32 out of 522 (6.1%) of the cases, respectively. However, the initial values of the disappeared TBR and TBR2 were relatively small (<0.1%). In the recordings with glucose management indicator <53 mmol/mol the $\varepsilon_{EPB}$ was 9.6% for 14 days with 30% data loss.

**Conclusions:** With a maximum of 30% data loss in 14-day CGM recordings, there is minimal impact of missing data on the clinical interpretation of various glucose metrics.

**Trial Registration:** ClinicalTrials.gov NCT05584293; https://clinicaltrials.gov/study/NCT05584293

***(Interact J Med Res 2024;13:e50849)*** doi:10.2196/50849

**KEYWORDS**

continuous glucose monitoring; missing data; clinical decision-making; clinical targets; time below range; TBR; diabetes mellitus; data interpretation; clinical practice; data analysis; continuous glucose monitoring metrics; glucose; diabetes; diabetic; metrics; data loss; decision-making; decision support; missing values; data science

## Introduction

Improved glucose sensing techniques have led to the increased availability of continuous glucose monitoring (CGM) technology for patients with diabetes. These minimally invasive sensors measure the glucose concentration of the interstitial fluid every 1 or 5 minutes, representing the blood glucose concentration with an average delay of 10-12 minutes [1,2]. CGM provides insights into glucose concentration fluctuations throughout the day and enables comparisons over time using predefined glucose metrics. This offers better prevention of out-of-range values compared to glycated hemoglobin ($HbA_{1c}$) measurements or traditional finger prick methods. Consequently, CGM has the potential to improve glycemic control and adherence to lifestyle and drug regimens [3-7]. Therefore, CGM devices have been included in clinical guidelines and standards of care for patients with type 1 and 2 diabetes [8].

Commonly used glucose metrics to assess glycemic control in clinical practice from CGM devices include time in range (TIR), representing the percentage of time spent within the glucose range of 70-180 mg/dL (3.9-10.0 mmol/L), time below range (TBR) in 2 levels, time above range (TAR) in 2 levels, glucose management indicator (GMI), and the coefficient of variation (CV). Although treatment targets are individualized, the general aim is to achieve a TIR of over 70% while minimizing TBR below 4% [9,10].

Incomplete data collection during CGM monitoring can affect the clinical interpretation of glucose metrics. Several factors such as loss of connectivity, sensor or reader malfunction, depleted battery, or delayed interaction with the device when data are only temporarily stored on the sensor, could lead to data loss. This data loss can introduce bias in the estimation of glucose metrics [11]. Evaluating the potential impact of data loss on clinical interpretations and determining acceptable levels of error is crucial for clinical decision-making [12].

The type and distribution characteristics of missing data are important when evaluating data loss. There are 3 types of missing data, they are, missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) [13-15]. MCAR implies no systematic differences between participants with complete and incomplete data. MAR occurs when missing values are independent of the missing variables, but the pattern of missing data is dependent on time. In the case of MNAR, valuable information is lost from the data and there is no general method to manage the missing data properly [16,17]. Consecutive missing data, or gaps, have a certain gap size and incidence in the data, which can be represented by a gap probability distribution. Missing data are characterized as MCAR when the gap probability follows an exponential decline, with minimal affecting analysis and outcomes. However, MAR patterns skew the research outcomes, as missing data are more prevalent at certain times than others. Therefore, insight into the gap probability distribution is essential for understanding the influence of missing data on desired outcomes [15,18-20].

Currently, the recommendation for reliable interpretation of CGM data is to evaluate either 10 consecutive days without data loss or 14 consecutive days with a maximum data loss of 30% [8,10,21]. However, these studies have not investigated the impact of deviations from missing data in common CGM metrics on clinical decisions for individual patients [12,22,23]. Furthermore, previous research has determined that at least 14-15 days of CGM data provide a good estimation of CGM metrics compared to monitoring every 3 months or $HbA_{1c}$ [21,24]. These studies primarily focused on the correlation between CGM data and $HbA_{1c}$, which does not give insight into the impact of missing data on the clinical interpretation of different CGM metrics. Additionally, it only reflects long-term glycemic control, overlooking the potential for assessing short-term variations afforded by CGM data. Therefore, this study aims to investigate the effects of data loss on glucose metrics in individual patient recordings and its influence on clinical decision-making.

## Methods

### Patient Inclusion and Data Collection

This study was performed in the Diabase cohort, which is a registry of adult patients with type 1 diabetes and those with type 2 diabetes who use CGM technology as part of their care. They are treated in Ziekenhuisgroep Twente (ZGT), a local hospital in the Netherlands (NCT05584293). The exclusion criteria were dependency on hemodialysis or inability to provide informed consent.

All patients included in the Diabase cohort between September 2020 and March 2022 were reported upon, which contains retrospective data from June 2016.

### Ethical Considerations

The study was performed in accordance with the Declaration of Helsinki, the guidelines of good clinical practice. The Medical Research Ethics Committees United (MEC-U) in Nieuwegein, the Netherlands (registration AW23.009/W20.197), reviewed and approved the protocol. Prior to participation, patients provided informed consent to collect their (retrospective) glucose sensors and to retrieve relevant patient information from electronic patient files (age, gender, $HbA_{1c}$, and BMI).

For this study, deidentified data were provided by the Diabase cohort, ensuring the confidentiality and privacy of participant information. Participants did not receive financial compensation for their participation in this study, as it solely involved the collection of data readily available from standard practice.

### CGM Derived Clinical Metrics

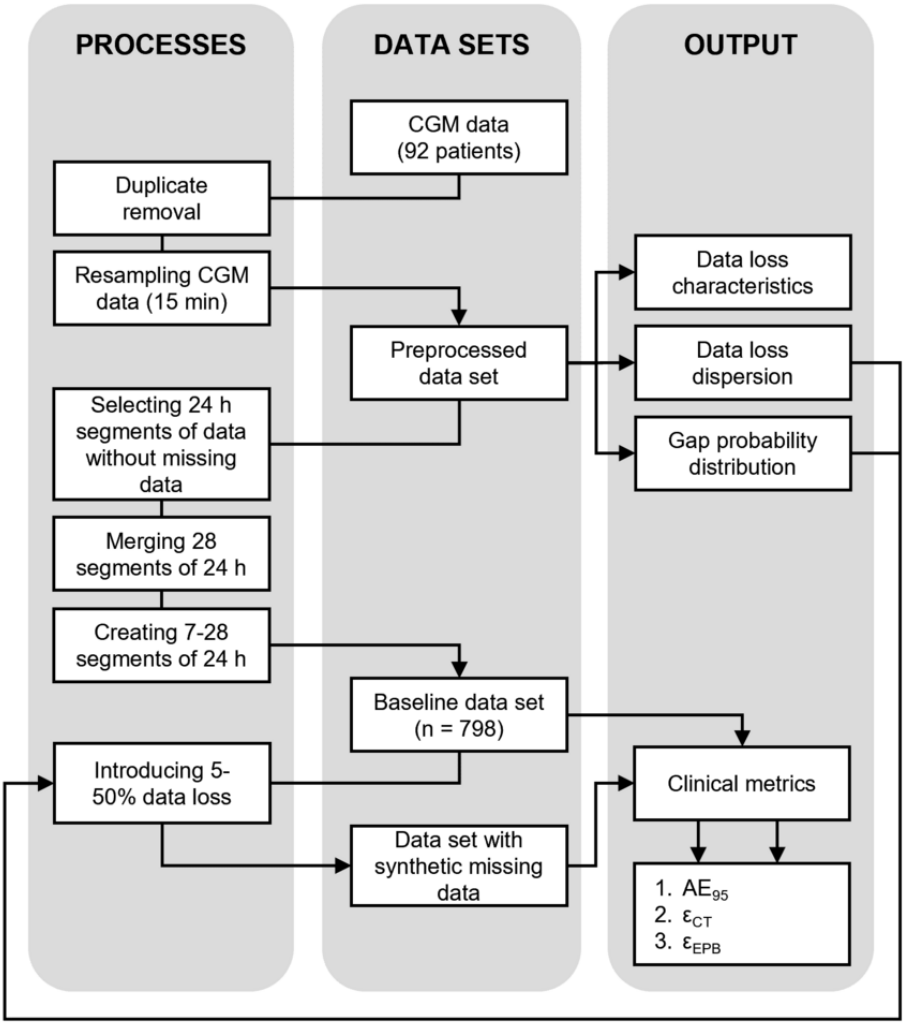
We collected CGM data using FreeStyle Libre sensors, and the data were stored in LibreView (Abbott Diabetes Care). The sensors have a storage capacity of up to 8 hours of data and measure for 2 weeks.

Derived from the CGM data were the mean glucose, TIR (glucose between 70 and 180 mg/dL or 3.9 and 10.0 mmol/L; %), TBR (glucose <70 mg/dL or 3.9 mmol/L; %), TBR level 2 (TBR2; glucose <54 mg/dL or 3.0 mmol/L; %), TAR (glucose >180 mg/dL or 10.0 mmol/L; %), TAR level 2 (TAR2; glucose >250.2 mg/dL or 13.9 mmol/L; %), SD glucose (mmol/L), CV (%), GMI (mmol/mol), low blood glucose index (LBGI), high blood glucose index (HBGI), and the risk index (RI; Multimedia Appendix 1) [8,10,25-27].

### Data Processing and Analysis

Data from LibreView were extracted as a CSV file and processed in Python 3.9 (extension Spyder 5.3.1; Python Software Foundation). Figure 1 illustrates the data processing steps from the original CGM data into the various outputs used in this study. Data from the FreeStyle Libre sensor were recorded at intervals of 13-19 minutes (with an average every 15 minutes). First, duplicate data and data from multiple sources within 14 minutes were removed from the data set. Subsequently, data were resampled to 1 sample per 15 minutes using linear interpolation, starting at midnight. Interpolated data points were marked as missing when the time difference between the 2 nearest original data points exceeded 19 minutes. These steps resulted in the preprocessed data set.

**Figure 1.** Data processing pipeline showing the data processing steps, the different data sets and the outputs. $AE_{95}$: 95th percentile of absolute error; CGM: continuous glucose monitoring; $\varepsilon_{CT}$: error clinical target; $\varepsilon_{EPB}$: error expert panel boundaries.




From the preprocessed data set, the data loss characteristics were determined (illustrated in Figure 1). The percentage of missing data was calculated, and we determined the percentage of time that patients adhered to the data loss guidelines outlined in prior studies. These studies indicated that maintaining >70% of CGM data over the last 14 days or 10 days out of the 14 days strongly correlates with mean glucose, estimated $HbA_{1c}$, TIR, and hyperglycemic metrics over a 3-month period. Consequently, our study adopts a 14-day measurement window with <30% data loss or a 10-day window with no data loss [8,21,24,28] The gap length and gap probability distribution were computed (Figure 1). To assess whether the missing data were MAR, we tested the fit of an exponential probability function to the gap probability distribution for gaps smaller than 96 samples [13]. The data loss dispersion over time was

researched by comparing the average data loss per patient of the preprocessed CGM data set between various time periods—hours within the day, days in the week, business days (Monday to Friday) versus weekends, months, seasons, and days of the year.

It is important to consider the circadian rhythm of glucose metabolism during data analysis. Numerous studies have shown diurnal variations in glucose tolerance, insulin secretion, and peripheral insulin sensitivity, with poorer glycemic control observed in the evening and at night in healthy individuals [29,30]. To account for circadian rhythm, we constructed a baseline data set for each patient by combining 28 segments of 24-hour data without missing data (Figure 1). In cases that were where available, we merged multiple periods of 28 days per patient for further analysis, excluding patients with less than 28 segments of data. From the 28-day recordings, we created shorter recordings ranging from 7 up to 28 days.

Hereafter, data loss ranging from 5% to 50% was randomly introduced into the baseline data set in accordance with the gap distribution and data loss dispersion determined from the preprocessed data set. This means the gap distribution of the synthetic missing data is as close as possible to the gap distribution found over the whole population in the preprocessed data set. Also, synthetic data have higher chance of missing following the found data loss dispersion over time. Therefore, the synthetic missing data mimics the missing data as seen in day-to-day patient care.

### Absolute Errors

The CGM metrics were calculated for each recording in the baseline and synthetic missing data set, and compared for every recording. The absolute error (AE), the median AE (MedAE), and the 95th percentile of the AE ($AE_{95}$) as a result of missing data were calculated for all CGM metrics (Figure 1).

### Clinical Target Errors ($\varepsilon_{CT}$)

The clinical targets follow clinical guidelines—<4% TBR, <1% TBR2, >70% TIR, <25% TAR, <5% TAR2, <36% CV, and <53 mmol/mol GMI [9,10]. The percentage of recordings that surpassed the clinical target cut-off because of missing data was determined as the clinical target error ($\varepsilon_{CT}$). This step is illustrated in Figure 1.

### Expert Panel Boundary Error ($\varepsilon_{EPB}$)

A panel of experts consisting of a diabetes specialist nurse, a diabetes nurse, and a technical physician (a medical specialist in diabetes-related technology in health care) was interviewed. Each expert was interviewed individually to discuss clinically relevant changes in CGM metrics. During the interviews, CGM metrics TIR, TAR, TAR2, TBR, TBR2, GMI, and CV were discussed separately. The experts were instructed to consider a generic patient with diabetes and to evaluate each CGM metric separately. They were asked to identify when a change in the CGM metric would likely result in therapy alteration, indicating clinical relevance. A clinically relevant change can be dependent on the initial value of the CGM metric. For example, a change in TBR from 2% to 4% can have more impact than a change from 8% to 10%, even though the absolute change is the same. Therefore, the experts determined clinically relevant changes for various initial values of all CGM metrics (Multimedia Appendix 2). From these discussions, the strictest relevant change per CGM metric was selected as the expert panel boundary (Multimedia Appendix 2). The percentage of recordings that exceed the defined expert panel boundary due to data loss is defined as the expert panel boundary error ($\varepsilon_{EPB}$), this step is depicted in Figure 1. When data loss resulted in an $\varepsilon_{CT}$ or $\varepsilon_{EPB}$ of more than 5%, we assumed that data loss had considerable influence on the CGM metric and should be interpreted with caution, consistent with the clinically significant criterion for a 5% increase in TIR [8].

### GMI Subgroup Analysis

To see whether missing data had a different influence on TBR, the recordings were divided into 3 GMI-based groups (<53, 53-64, and ≥64 mmol/mol), based on 3 commonly used $HbA_{1c}$ categories as determined by the American Diabetes Association [31]. Hereafter, the MedAE, $\varepsilon_{CT}$, and $\varepsilon_{EPB}$ were calculated for TBR in the 3 GMI groups.

### Statistical Analysis

The median and IQR were calculated for the clinical and sensor data characteristics, that is, the recording length, data loss, and overall CGM metrics.

The fit between the gap distribution and the exponential probability density function was tested with a 2-sided Kolmogorov-Smirnov test and the sum of the squared error. Differences in data loss dispersion between several time periods were evaluated using the Kruskal-Wallis test with post hoc Dunn test and Bonferroni correction. Differences between the glucose metrics between the preprocessed data set and the baseline data set with 28 days were evaluated using Kruskal-Wallis test with Bonferroni correction. For these statistical analyses, the packages SciPy and scikit-posthocs in Python were used [32,33]. All data visualizations were made with the Python package Matplotlib [34].

Scatterplots were generated for CGM metrics from 14-day recordings with complete data versus those with 30% synthetic missing data. Additionally, histograms of the AE were constructed from 14-day recordings with 30% data loss in TBR and TBR2. The histogram bin widths were determined with Freedman-Diaconis rule. A P value <.01 was considered significant.

## Results

### Data and Patient Characteristics

Between September 2020 and March 2022, 92 patients using a FreeStyle Libre CGM device were included in the Diabase study. Descriptive statistics of the population can be found in Table 1.

**Table 1.** Characteristics of the population and their data used in this research.

| | Values |
|---|---|
| Type 1 diabetes, n (%) | 79 (95.8) |
| Type 2 diabetes, n (%) | 13 (14.2) |
| Men, n (%) | 47 (51) |
| Age (years), median (IQR) | 52 (37.3-60.5) |
| BMI (kg/m$^2$), median (IQR) | 26.2 (23.1-29.4) |
| $HbA_{1c}$[a] (%), median (IQR) | 7.6 (7.0-8.3) |
| $HbA_{1c}$ (mmol/mol), median (IQR) | 60 (53.3-67.0) |
| Recording length (days), median (IQR) | 655 (499-925) |
| Data loss (%), median (IQR) | 13.5 (6.3-35.6) |
| Time recordings had 0% data loss for 10 days (%), median (IQR) | 0.0 (0.0-0.3) |
| Time recordings had <30% data loss for 14 days (%), median (IQR) | 88.9 (63.0-98.7) |

[a]$HbA_{1c}$: glycated hemoglobin.

## Data Loss Statistics

The gap probability distribution of the preprocessed data set differs significantly from a fitted exponential probability density function (P=.002). This means the data loss is not completely random (Multimedia Appendix 3).

During the hours of 10-11 PM, 11 PM to midnight, and midnight to 1 AM, a higher data loss of 35.8%, 43.3%, and 35.3%, respectively, was observed compared to other hours of the day with an average data loss of 10.4% (IQR 2.7%-30.5%; P<.001). Therefore, to mimic the unequal dispersion we created synthetic missing data following the probability of missing data during the day. No significant differences in data loss were found between other time periods.

## Baseline Data Set

Among the 92 patients, 84 (91%) patients had sufficient data to construct a data set of 28 segments with complete 24-hour segments of data. These 84 patients contributed to a total of 798 recordings over 28 days, forming the baseline data set. Each patient provided median of 7.0 (IQR 3.8-14.0) in 28-day recordings. The median (IQR) for TIR, TBR, TBR2, TAR, and TAR2 of all 28-day recordings were 60.3% (53.2%-67.8%), 3.3% (1.4%-5.8%), 0.3% (0.0%-1.0%), 35.4% (27.4%-43.3%), and 9.4% (5.5%-13.9%), respectively. No significant differences were found between all the clinical metrics of the preprocessed data and the 28-day baseline data set, indicating that the selected baseline data set reflects the original data.

## Expert Panel

The experts unanimously agreed that the most important metrics were TBR and TBR2, resulting in strict expert panel boundaries. In the TBR range of 0%-4% and the TBR2 range of 0%-5%, a difference of 1% was deemed clinically relevant (Multimedia Appendix 2). However, as the TBR increased, the experts allowed for higher changes, accepting a maximum of 5% for TBR and 10% for TBR2 when the initial value was 100%. Expert panel boundaries for the remaining glucose metrics can be found in Multimedia Appendix 2.

## CGM Metrics Without and Those With Synthetic Data Loss

In the baseline data set, 734 (92.0%), 408 (51.1%), 796 (99.8%), and 767 (96.1%) of the total (N=798) 7-day recordings contained TBR, TBR2, TAR, and TAR2, respectively. For 28-day recordings, this is 798 (100%), 611 (76.6%), 798 (100%), and 790 (99.0%) for TBR, TBR2, TAR, and TAR2, respectively.

The scatterplots of Figure 2A represent the relation between the original 14-day data of TBR and TBR2, and with 30% data loss, per individual recording. These figures show that 30% data loss in these metrics results in small deviations from the true value with data loss, as all data points are close to the identity line. More recordings fell into the $\varepsilon_{EPB}$ area for TBR, as the expert panel boundary was stricter compared to TBR2. The histograms of Figure 2B show the AE, MedAE, and $AE_{95}$ for these CGM metrics as a result of 30% data loss in a 14-day recording. The MedAE in the histograms is small, indicating that the majority of errors are small. However, the histograms also display a long tail, indicating that there are instances where the AE is larger than the MedAE.

Figure 3 shows the values of $AE_{95}$, $\varepsilon_{CT}$, and $\varepsilon_{EPB}$ for the CGM metrics TBR and TBR2, with missing data increasing from 5% to 50% over a period of 7 to 28 days. This analysis includes only the recordings that had TBR episodes. As expected, the $AE_{95}$, $\varepsilon_{CT}$, and $\varepsilon_{EPB}$ increase as the percentage of data loss increases. When more days are available, the influence of data loss is reduced (Figure 3A). Applying the current guidelines of 30% data loss for a 14-day recording period, we observed an $AE_{95}$ of 1.0% for TBR and 0.5% for TBR2. The corresponding $\varepsilon_{CT}$ values were 29 out of 786 (3.7%) recordings for TBR and 28 out of 522 (5.4%) recordings for TBR2 and the $\varepsilon_{EPB}$ were 28 out of 786 recordings (5.0%) for TBR 1 out of 522 recordings and 0.2% for TBR2. These findings suggest that the impact of missing data on CGM metrics can vary significantly depending on the specific metric and recording period used.

**Figure 2.** (A) The density scatterplots of the TBR and (B) TBR2 indicates the relation between the 14-day-long original data without missing data (horizontal axis) and the data with 30% data loss (vertical axis). More detail of the critical areas of TBR and TBR2 are illustrated in (C) and (D). The color bar indicates the Gaussian kernel-density estimate of the recordings. The black dashed lines represent the expert panel boundaries, and the horizontal and vertical black dotted lines represent the clinical targets. Values falling outside the clinical target and expert panel boundary, in the hatched areas, are labeled as errors ($\varepsilon_{CT}$ and $\varepsilon_{EPB}$). (E) The histograms show the AE (%) with the median AE (MedAE) and $AE_{95}$ as a result of 30% data loss of the TBR and (F) TBR2. AE: absolute error; $AE_{95}$: 95th percentile of AE; CT: clinical target; EPB: expert panel boundary; MedAE: median absolure error; TBR: time below range; TBR2: time below range level 2; $\varepsilon_{CT}$: clinical target errors; $\varepsilon_{EPB}$: expert panel boundary errors.

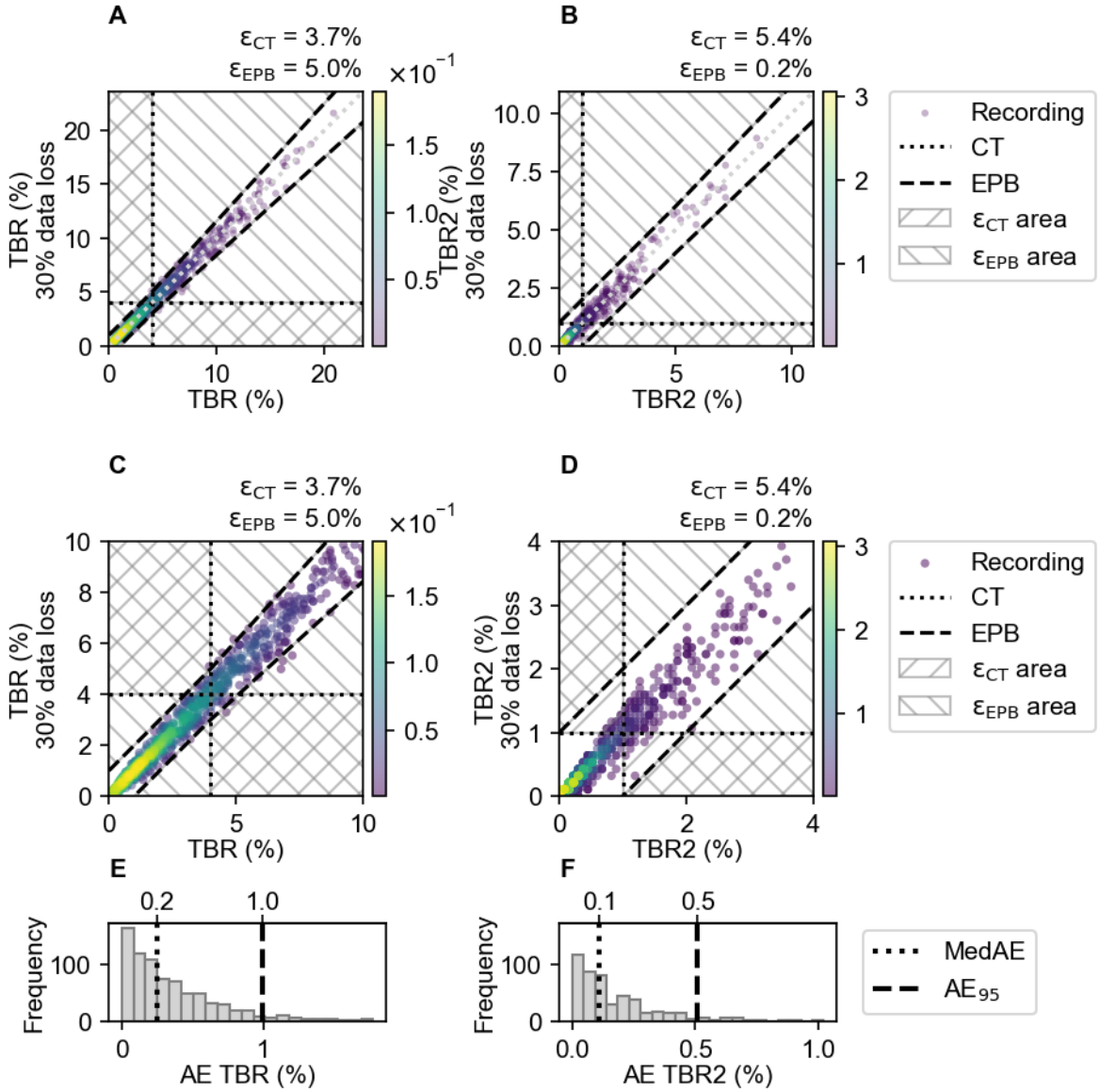
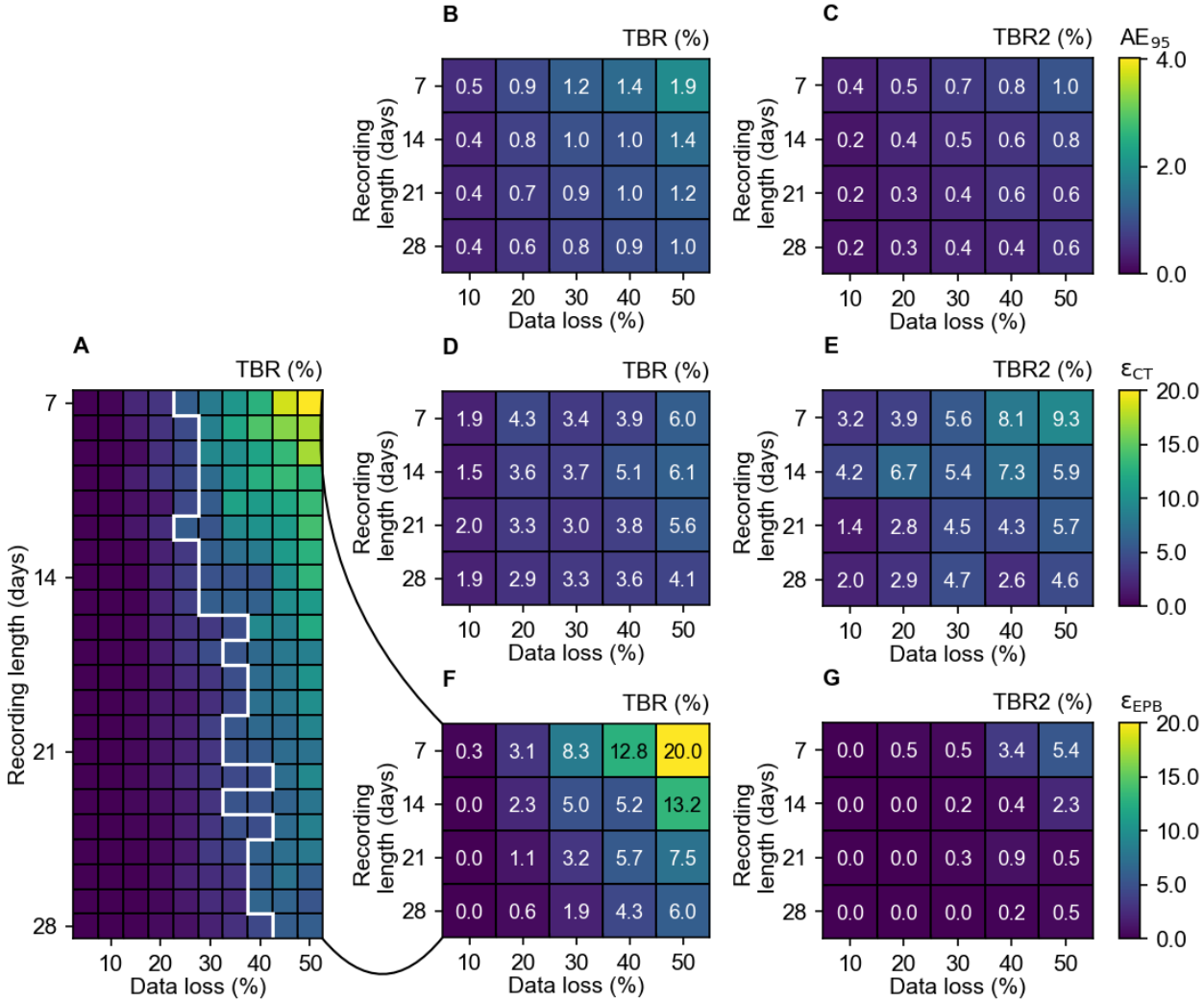


**Figure 3.** (A) The $\varepsilon_{EPB}$ of 5%-50% data loss in recordings (horizontal axis) of 7-28 days (vertical axis) of TBR. The white line indicates the boundary where the $\varepsilon_{EPB}$ exceeds 5%. (B,C) show the $AE_{95}$, (D,E) the clinical target errors ($\varepsilon_{CT}$), and (F,G) the $\varepsilon_{EPB}$ of TBR and TBR2 respectively of 10%, 20%, 30%, 40%, and 50% data loss in recordings of 7, 14, 21, and 28 days. The color bar indicates the errors in percentages. $AE_{95}$: 95th percentile of the absolute error; TBR: time below range; TBR2: time below range level 2; $\varepsilon_{EPB}$: expert panel boundary errors; $\varepsilon_{CT}$: clinical target errors.




For a 14-day measurement period with 30% data loss, the $AE_{95}$ was 2.4%, 1.3%, and 1.0%, $\varepsilon_{CT}$ was 17 (2.1%), 23 (2.9%), and 14 (1.8%) and $\varepsilon_{EPB}$ was 0 (0.0%), 0 (0.0%), and 0 (0.0%) for TIR, CV, and GMI, respectively (Figures S1 and S2 in Multimedia Appendix 4). For TAR and TAR2 the $\varepsilon_{EPB}$ was 0.0% for all cases (Figure S3 in Multimedia Appendix 4).

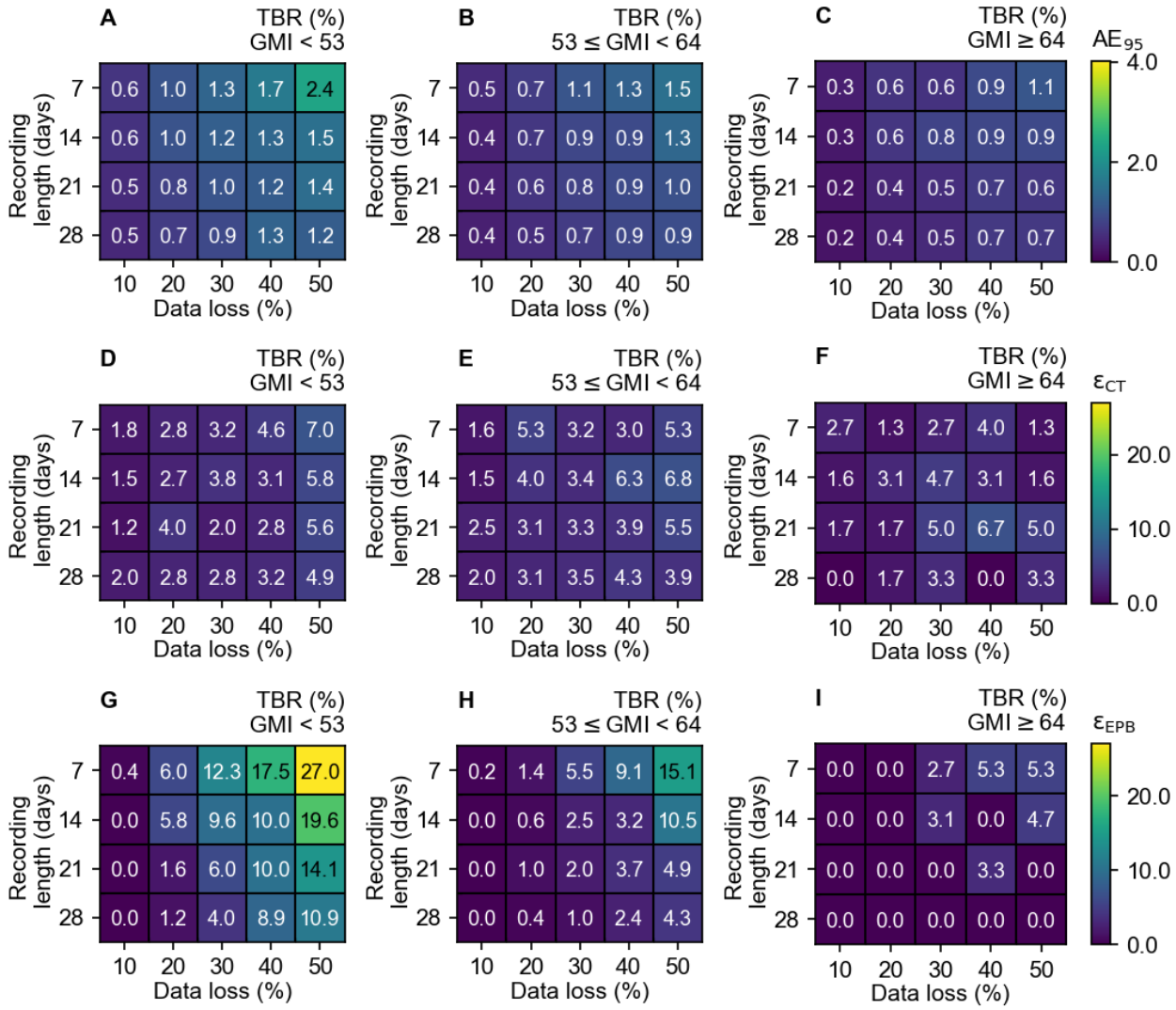
TBR and TBR2 have the highest errors, especially with greater data loss in shorter recordings, with a maximum $\varepsilon_{EPB}$ of 147 out of 736 (20.0%) of the recordings for TBR and 22 out of 408 (5.4%) of the recordings for TBR2, compared to a maximum of 5 out of 798 (0.6%) for TIR, CV, and GMI. In 2 (0.3%) of the 14-day recordings, the TBR value went from 0.2% to 0.0%, resulting in its disappearance due to data loss. Similarly, the TBR2 metric disappeared in 32 (6.1%) recordings in the 14-day recordings with 30% data loss. The median TBR2 value of the original recordings was 0.07% (IQR 0.07%-0.15%). No other metrics disappeared due to data loss.

For the SD, LBGI, HBGI, and RI, only the $AE_{95}$ was available. The overall $AE_{95}$ was low, with 0.1 for SD, 0.2 for LBGI, and 0.6 for HBGI and RI. The highest $AE_{95}$ of 0.8% was observed for the HBGI and RI metrics in a 7-day recording with 50% data loss (Figure S4 in Multimedia Appendix 4).

## GMI Subgroup Analysis

The recordings were divided into 260 (32.6%) low GMI (<54 mmol/mol), 445 (55.8%) moderately elevated (53-64 mmol/mol), and 93 (11.7%) elevated GMI recordings (≥64 mmol/mol). Figure 4 shows the values of $AE_{95}$, $\varepsilon_{CT}$, and $\varepsilon_{EPB}$ for TBR, with missing data increasing from 5% to 50% over a period of 7 to 28 days for the GMI subgroups. The error caused by missing data is highest in the low GMI group (Figures 4A, 4D, and 4G). The $AE_{95}$, $\varepsilon_{CT}$, and $\varepsilon_{EPB}$ are 1.2%, 10 (3.8%), and 25 out of recordings 260 (9.6%), respectively, for a recording of 14 days with 30% data loss. This means that for 9.6% of the recordings a clinical expert would see a relevant change.

**Figure 4.** (A-C) The $AE_{95}$ are shown, (D-F) the $\varepsilon_{CT}$, and (G-I) the $\varepsilon_{EPB}$ of TBR within 3 GMI groups of <53 mmol/mol, 53 mmol/mol≤ GMI <64 mmol/mol, and GMI ≥64 mmol/mol. These panels illustrate the impact of 10%, 20%, 30%, 40%, and 50% data loss in recordings of 7, 14, 21, and 28 days. The color bar indicates the errors in percentages. $AE_{95}$: 95th percentile of the absolute error; GMI: glucose management indicator; TBR: time below range; $\varepsilon_{CT}$: clinical target errors; $\varepsilon_{EPB}$: expert panel boundary errors.




## Discussion

### Principal Findings

This study provides a thorough analysis of missing data's impact on real-world CGM recordings for patients with diabetes. By merging data analysis and expert evaluations, it assesses the clinical implications of missing data on CGM metrics, improving our understanding of its practical effects and challenges. Findings indicate that a minimum of 14 days of glucose data collection with no more than 30% missing data suffice for clinical decision-making, ensuring adequate patient care.

The key finding of this study is that the interpretation of TBR metrics is more vulnerable to greater clinical consequences due to missing data compared to other metrics. The results indicate that up to 30% data loss in a 14-day recording results in a misinterpretation of the glucose metric in 5.0% of the time. Therefore, it should be realized that 30% data loss over a 14-day measurement will occasionally lead to false clinical interpretations of TBR. Furthermore, when analyzing the data by GMI levels, it is notable that in 9.6% of recordings featuring low GMI levels, a clinical expert would observe a relevant change due to missing data on TBR. These findings underscore the substantial impact of missing data on TBR, emphasizing the importance of cautious interpretation in clinical practice.

Similarly, for TBR2, 14-day recordings with 30% data loss resulted in misinterpretation in 0.2% of cases, and the complete disappearance of TBR2 episodes occurred in 6.1% of the cases. One might, therefore, suggest that severe hypoglycemic episodes would be missed, which can have potentially serious clinical consequences as symptoms might not always be recognized when they occur during nighttime [35]. However, this complete loss of TBR2 episodes occurs only when the actual TBR2 value is below the clinical target of 1%, and thus, would not have required any action. In contrast, missing data had limited influence on the clinical interpretation of other glucose metrics such as TIR, CV, GMI, TAR and TAR2. This limited influence was partly caused by broader expert panel boundaries for these

metrics compared to TBR and TBR2. The expert panel accepted errors of 1%-5% for TBR2 and 1%-10% for TBR, while the accepted errors for TIR, TAR, and TAR2 ranged between 5% and 10%. When TBR is not a primary concern, 14 days of CGM data can be interpreted with 50% missing data. Even measurements shorter than 14 days could be used, but the representativeness of the data for long-term glucose values will diminish, and the correlation of the metrics with $HbA_{1c}$ may decrease [21].

Based on the findings of this study, we propose several recommendations. First, we support the current recommendation of having a minimum of 70% data available in a CGM recording of 14 days. However, health care practitioners should be aware that there is still a chance of misinterpreting TBR. The guideline of a 10-day CGM recording with no data loss might be impractical, as the real-life CGM data analyzed in this study rarely met this criterion [8]. Therefore, our second recommendation is to use a measurement period of at least 14 days, which aligns with the durability of the current sensors. However, this result should be validated by other studies with different glucose sensors.

In this study, the expert panel boundaries are an important feature that may contribute to current clinical guidelines as they give insight into when CGM metrics result in clinically relevant changes due to data loss. The current clinical targets for most CGM metrics are valuable in the clinical setting to serve as goals for optimal glucose management. However, the targets are a limited tool for determining when a patient has a clinically relevant improvement or deterioration of their glucose management. When assessing the clinical consequences, the clinical target may not always be useful because significant changes due to data loss may not result in a different categorization if the initial value is already far from the target. However, such changes may still be clinically relevant.

In the literature, the CGM metrics SD, LBGI, HBGI, and RI are defined [21]. However, there are currently no targets defined for these metrics. Consequently, the expert panel was unable to determine such a boundary for these metrics. Nevertheless, $AE_{95}$ can provide insight into the error size, but not indicate the impact on clinical decision-making. Previous studies investigating the influence of missing data on glycemic variability metrics reported small MedAEs [12,22]. This corresponds to our findings of small MedAEs, suggesting that even with excessive data loss, the CGM metrics would not be altered significantly. However, the $AE_{95}$ is in some cases relatively high, indicating that in 5% of the population data loss can have quite significant clinical consequences. Therefore, we reported on $AE_{95}$ instead of MedAE in this study.

The data loss in the real-world CGM data was MNAR, as the gap probability distribution did not follow an exponential function. Next to that, the chance of missing data during the night (10 PM-1 AM) was higher. MNAR data loss has a profound impact on the effects of missing data and synthetic data loss cannot be applied arbitrarily. Therefore, in this research, we mimicked these characteristics of data loss to create a simulation that closely resembled the actual missing data experienced by patients using CGM devices in everyday life.

## Limitations

Several potential limitations of this study need to be addressed. In clinical practice, all CGM metrics are typically evaluated together rather than individually which was done to determine the clinical expert boundaries. Additionally, these metrics will be reviewed at every clinical visit, further reducing the actual risk that missing data will cause a change in treatment. Another potential limitation of our study is the small number of experts involved in defining the boundaries for CGM metrics. The subjectivity and potential biases inherent in expert input can pose challenges. Furthermore, the cut-off to accept an $\varepsilon_{EPB}$ at 5% was chosen arbitrarily and choosing a different cut-off could influence the interpretation of the results. It is important to note that defining different boundaries than those presented in this study could influence the observed errors.

It should also be considered that this research has been performed on data measured only with the FreeStyle Libre 1 sensor, and some findings might differ for other devices. For example, the 8-hour data storage capacity may explain why more data are lost during the early night (10 PM-1 AM), which is a critical time for glucose measurements [36]. However, the presented research methodology can be easily adopted and applied to other CGM devices or populations. The results presented in our study were obtained from patients with diabetes, predominantly type 1 diabetes, who were treated in a hospital in the Netherlands. Differences in patient demographics, disease progression, and device accuracy could impact the applicability of findings across diverse populations. Therefore, caution should be exercised when extrapolating these results to other populations. However, the average $HbA_{1c}$ value of 7.6% (IQR 7.0%-8.3%; 60.0 mmol/mol, IQR 53.3-67.0 mmol/mol) was comparable with a large cohort from Germany and Austria, with a mean $HbA_{1c}$ of 7.8% (IQR 6.9%-8.9%; 62, IQR 52.0-74.0 mmol/mol), suggesting some level of generalizability [37]. Furthermore, the presented methodology can be applied to investigate the consequences of missing data in other diabetes populations using CGM.

## Methodological Decisions

Some noteworthy methodological decisions were made in this study. First, we decided to include all the available 24-hour windows, thus including multiple 28-day recordings per patient in the data set. With this approach, there were no significant differences in CGM metrics between the preprocessed data with the baseline data set, validating the inclusion of multiple recordings. Second, not all patients used their CGM device continuously. Patients may have periods of several months where they did not use a CGM device. We marked this as missing data, which may have led to a potential overestimation of the reported data loss, compared to day-to-day CGM use. Third, we decided to create recordings of 7 to 28 days to study a range of CGM measurement lengths. The studies of Riddlesworth et al [21] and Xing et al [24] suggest that at least 14-15 days of CGM data provide a good estimation of CGM metrics compared to monitoring every 3 months or $HbA_{1c}$. Also,

Akturk et al [23] state that while the optimal recording length depends on the size of the gaps, duration of 14 days generally proves to be adequate. Using measurements of 7-28 days, we can ensure that an adequate measurement duration was covered.

### Future Research

Future research should focus on including a heterogeneous population of patients with type 1, type 2, and other subtypes of diabetes. Next to that, more commonly used sensors should be added to the analysis to give a more generalizable result. Finally, the expert panel should be expanded and implemented similarly to the Parkes and Clarke error grids, involving a larger and more diverse panel of experts, to enhance the reliability and generalizability of the established boundaries [38,39].

### Conclusions

To conclude, our aim was to examine the impact of data loss on glucose metrics within individual patient recordings from continuous glucose monitors and its implications for clinical decision-making. Through integrated data analysis and expert evaluations, we underscore the importance of comprehending missing data's clinical consequences and recommend a maximum of 30% missing data in 14-day CGM recordings to enhance accurate interpretation and glucose management, acknowledging the possibility of misinterpreting TBR even with this threshold. For reliable interpretation of TBR in recordings with a low GMI, data loss should be below10%. Further research is needed to explore the consequences of missing data in diverse populations using various CGM devices, emphasizing the importance of comprehensive data collection for optimal glucose management and clinical decision-making.

## Acknowledgments

The authors used the generative artificial intelligence tool ChatGPT 3.5 by OpenAI (2022) for language checking, suggestions on sentence structure, and improving overall readability. The authors thank Anouk Veldhuis, Ziekenhuisgroep Twente, for her contribution to Diabase. This research was financially supported by the grant Exceptional and Deep Intelligent Coach (EDIC) financed by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), grant (628.011.02). NdB and CIRB are appointed by this grant. Diabase was financially supported by an unrestricted grant from the Pioneers in Health Care Innovation Fund 2020 established by the University of Twente, Medisch Spectrum Twente, and Ziekenhuisgroep Twente.

## Data Availability

The data sets generated and analyzed during this study are not publicly available due to the ongoing inclusion of subjects but are available from the principal investigator (TU) on reasonable request.

## Authors' Contributions

Conceptualization, methodology, and validation were performed by NdB, CIRB, USY, MMRVH, and GDL. Formal analysis, visualization, and software were completed by NdB and CIRB. Writing, the original draft preparation was done by NdB and CIRB. Writing, the review and editing were performed by NdB, CIRB, USY, TU, MMRVH, GDL, PHV, and HJH. Supervision was performed by MMRVH, GDL, PHV, and HJH. All authors have read and agreed to the published version of the study. GDL is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Formulas of the clinical continuous glucose monitoring (CGM) metrics.
[DOCX File , 28 KB - ijmr_v13i1e50849_app1.docx ]

Multimedia Appendix 2
Expert panel boundaries and the clinically relevant change, using the initial metric value as a baseline, for the seven most commonly used clinical continuous glucose monitoring (CGM) metrics, along with their respective clinical targets.
[DOCX File , 24 KB - ijmr_v13i1e50849_app2.docx ]

Multimedia Appendix 3
Description of the method and results of fitting the gap probability distribution of preprocessed continuous glucose monitoring (CGM) data with the exponential and other common probability mass functions.
[DOCX File , 54 KB - ijmr_v13i1e50849_app3.docx ]

Multimedia Appendix 4

Density scatterplots for time in range (TIR) and coefficient of variation (CV) and the error plots for glucose metrics TIR, CV, glucose management indicator (GMI), time above range (TAR), TAR level 2 (TAR2), SD, low blood glucose index (LBGI), high blood glucose index (HBGI), and risk index (RI).
[DOCX File , 968 KB - ijmr_v13i1e50849_app4.docx ]

## References


1. Schmelzeisen-Redeker G, Schoemaker M, Kirchsteiger H, Freckmann G, Heinemann L, Del Re L. Time delay of CGM sensors: relevance, causes, and countermeasures. J Diabetes Sci Technol 2015;9(5):1006-1015 [FREE Full text] [doi: 10.1177/1932296815590154] [Medline: 26243773]
2. Zaharieva DP, Turksoy K, McGaugh SM, Pooni R, Vienneau T, Ly T, et al. Lag time remains with newer real-time continuous glucose monitoring technology during aerobic exercise in adults living with type 1 diabetes. Diabetes Technol Ther 2019;21(6):313-321 [FREE Full text] [doi: 10.1089/dia.2018.0364] [Medline: 31059282]
3. Jangam S, Dunn T, Xu Y, Hayter G, Ajjan RA. Flash glucose monitoring improves glycemia in higher risk patients: a longitudinal, observational study under real-life settings. BMJ Open Diabetes Res Care 2019;7(1):e000611 [FREE Full text] [doi: 10.1136/bmjdrc-2018-000611] [Medline: 31114698]
4. Dunn TC, Xu Y, Hayter G, Ajjan RA. Real-world flash glucose monitoring patterns and associations between self-monitoring frequency and glycaemic measures: a European analysis of over 60 million glucose tests. Diabetes Res Clin Pract 2018;137:37-46 [FREE Full text] [doi: 10.1016/j.diabres.2017.12.015] [Medline: 29278709]
5. Rodbard D. Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes. Diabetes Technol Ther 2017;19(S3):S25-S37 [FREE Full text] [doi: 10.1089/dia.2017.0035] [Medline: 28585879]
6. Taylor PJ, Thompson CH, Luscombe-Marsh ND, Wycherley TP, Wittert G, Brinkworth GD. Efficacy of real-time continuous glucose monitoring to improve effects of a prescriptive lifestyle intervention in type 2 diabetes: a pilot study. Diabetes Ther 2019;10(2):509-522 [FREE Full text] [doi: 10.1007/s13300-019-0572-z] [Medline: 30706365]
7. Oser TK, Cucuzzella M, Stasinopoulos M, Moncrief M, McCall A, Cox DJ. An innovative, paradigm-shifting lifestyle intervention to reduce glucose excursions with the use of continuous glucose monitoring to educate, motivate, and activate adults with newly diagnosed type 2 diabetes: pilot feasibility study. JMIR Diabetes 2022;7(1):e34465 [FREE Full text] [doi: 10.2196/34465] [Medline: 35050857]
8. Danne T, Nimri R, Battelino T, Bergenstal RM, Close KL, DeVries JH, et al. International consensus on use of continuous glucose monitoring. Diabetes Care 2017;40(12):1631-1640 [FREE Full text] [doi: 10.2337/dc17-1600] [Medline: 29162583]
9. Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care 2019;42(8):1593-1603 [FREE Full text] [doi: 10.2337/dci19-0028] [Medline: 31177185]
10. Battelino T, Alexander CM, Amiel SA, Arreaza-Rubin G, Beck RW, Bergenstal RM, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol 2023;11(1):42-57. [doi: 10.1016/S2213-8587(22)00319-9] [Medline: 36493795]
11. Rubin DB. Inference and missing data. Biometrika 1976;63(3):581-592. [doi: 10.1093/biomet/63.3.581]
12. Kucharski P, Pagacz K, Szadkowska A, Młynarski W, Romanowski A, Fendler W. Resistance to data loss of glycemic variability measurements in long-term continuous glucose monitoring. Diabetes Technol Ther 2018;20(12):833-842. [doi: 10.1089/dia.2018.0247] [Medline: 30403500]
13. Zhou N, Patel D, Iyengar A, Shrivastava S, Bhamidipaty A. A verifiable imputation analysis for univariate time series and enabling package. 2020 Presented at: Proceedings - 2020 IEEE International Conference on Big Data (Big Data); 2020 Dec 10; Atlanta, Georgia, United States p. 1764-1773. [doi: 10.1109/bigdata50022.2020.9377909]
14. Moritz S, Sardá A, Bartz-Beielstein M, Zaefferer M, Stork J. Comparison of different methods for univariate time series imputation in R. arXiv 2015 Oct 13:1-20 [FREE Full text]
15. Braem CIR, Yavuz US, Hermens HJ, Veltink PH. Missing data statistics provide causal insights into data loss in diabetes health monitoring by wearable sensors. Sensors (Basel) 2024;24(5):1526 [FREE Full text] [doi: 10.3390/s24051526] [Medline: 38475061]
16. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol 2013;64(5):402-406 [FREE Full text] [doi: 10.4097/kjae.2013.64.5.402] [Medline: 23741561]
17. García-Laencina PJ, Sancho-Gómez JL, Figueiras-Vidal AR. Pattern classification with missing data: a review. Neural Comput Appl 2009;19(2):263-282. [doi: 10.1007/s00521-009-0295-6]
18. Khayati M, Lerner A, Tymchenko Z, Cudré-Mauroux P. Mind the gap: an experimental evaluation of imputation of missing values techniques in time series. Proc VLDB Endow 2020;13(5):768-782. [doi: 10.14778/3377369.3377383]
19. Moritz S, Bartz-Beielstein T. ImputeTS: time series missing value imputation in R. R J 2017;9(1):207. [doi: 10.32614/rj-2017-009]
20. Blankers M, Koeter MWJ, Schippers GM. Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. J Med Internet Res 2010;12(5):e54 [FREE Full text] [doi: 10.2196/jmir.1448] [Medline: 21169167]

21. Riddlesworth TD, Beck RW, Gal RL, Connor CG, Bergenstal RM, Lee S, et al. Optimal sampling duration for continuous glucose monitoring to determine long-term glycemic control. Diabetes Technol Ther 2018;20(4):314-316. [doi: 10.1089/dia.2017.0455] [Medline: 29565197]
22. Kingsnorth AP, Whelan ME, Orme MW, Routen AC, Sherar LB, Esliger DW. Resistance to data loss from the Freestyle Libre: impact on glucose variability indices and recommendations for data analysis. Appl Physiol Nutr Metab 2021;46(2):148-154. [doi: 10.1139/apnm-2020-0386] [Medline: 32813987]
23. Akturk HK, Herrero P, Oliver N, Wise H, Eikermann E, Snell-Bergeon J, et al. Impact of different types of data loss on optimal continuous glucose monitoring sampling duration. Diabetes Technol Ther 2022;24(10):749-753. [doi: 10.1089/dia.2022.0093] [Medline: 35653736]
24. Xing D, Kollman C, Beck RW, Tamborlane WV, Laffel L, Buckingham BA, Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Optimal sampling intervals to assess long-term glycemic control using continuous glucose monitoring. Diabetes Technol Ther 2011;13(3):351-358 [FREE Full text] [doi: 10.1089/dia.2010.0156] [Medline: 21299401]
25. Bergenstal RM, Beck RW, Close KL, Grunberger G, Sacks DB, Kowalski A, et al. Glucose management indicator (GMI): a new term for estimating A1C from continuous glucose monitoring. Diabetes Care 2018;41(11):2275-2280 [FREE Full text] [doi: 10.2337/dc18-1581] [Medline: 30224348]
26. Kovatchev BP, Cox DJ, Kumar A, Gonder-Frederick L, Clarke WL. Algorithmic evaluation of metabolic control and risk of severe hypoglycemia in type 1 and type 2 diabetes using self-monitoring blood glucose data. Diabetes Technol Ther 2003;5(5):817-828. [doi: 10.1089/152091503322527021] [Medline: 14633347]
27. Fabris C, Patek SD, Breton MD. Are risk indices derived from CGM interchangeable with SMBG-based indices? J Diabetes Sci Technol 2015;10(1):50-59 [FREE Full text] [doi: 10.1177/1932296815599177] [Medline: 26275643]
28. Beck RW, Connor CG, Mullen DM, Wesley DM, Bergenstal RM. The fallacy of average: how using HbA alone to assess glycemic control can be misleading. Diabetes Care 2017;40(8):994-999 [FREE Full text] [doi: 10.2337/dc17-0636] [Medline: 28733374]
29. Poggiogalle E, Jamshed H, Peterson CM. Circadian regulation of glucose, lipid, and energy metabolism in humans. Metabolism 2018;84:11-27 [FREE Full text] [doi: 10.1016/j.metabol.2017.11.017] [Medline: 29195759]
30. Martorina W, Tavares A. Effects of melatonin on glycemic variability in type 2 diabetes mellitus: protocol for a crossover, double-blind, placebo-controlled trial. JMIR Res Protoc 2023;12:e47887 [FREE Full text] [doi: 10.2196/47887] [Medline: 37410852]
31. American Diabetes Association Professional Practice Committee. 6. Glycemic targets: standards of medical care in diabetes-2022. Diabetes Care 2022;45(Suppl 1):S83-S96. [doi: 10.2337/dc22-S006] [Medline: 34964868]
32. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Polat, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
33. Terpilowski MA. scikit-posthocs: pairwise multiple comparison tests in Python. J Open Source Softw 2019;4(36):1169. [doi: 10.21105/joss.01169]
34. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007;9(3):90-95. [doi: 10.1109/mcse.2007.55]
35. Fidler C, Elmelund Christensen T, Gillard S. Hypoglycemia: an overview of fear of hypoglycemia, quality-of-life, and impact on costs. J Med Econ 2011;14(5):646-655 [FREE Full text] [doi: 10.3111/13696998.2011.610852] [Medline: 21854191]
36. den Braber N, Vollenbroek-Hutten MMR, Westerik KM, Bakker SJL, Navis G, van Beijnum BF, et al. Glucose regulation beyond HbA1C in type 2 diabetes treated with insulin: real-world evidence from the DIALECT-2 cohort. Diabetes Care 2021;44(10):2238-2244 [FREE Full text] [doi: 10.2337/dc20-2241] [Medline: 34301732]
37. Gerstl EM, Rabl W, Rosenbauer J, Gröbe H, Hofer SE, Krause U, et al. Metabolic control as reflected by HbA1c in children, adolescents and young adults with type-1 diabetes mellitus: combined longitudinal analysis including 27,035 patients from 207 centers in Germany and Austria during the last decade. Eur J Pediatr 2008;167(4):447-453. [doi: 10.1007/s00431-007-0586-9] [Medline: 17924142]
38. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987;10(5):622-628. [doi: 10.2337/diacare.10.5.622] [Medline: 3677983]
39. Parkes JL, Slatin SL, Pardo S, Ginsberg BH. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. Diabetes Care 2000;23(8):1143-1148. [doi: 10.2337/diacare.23.8.1143] [Medline: 10937512]


## Abbreviations

**$\varepsilon_{CT}$:** clinical target error
**$\varepsilon_{EPB}$:** expert panel boundary error
**AE:** absolute error
**$AE_{95}$:** 95th percentile of the absolute error

**CGM:** continuous glucose monitoring
**CV:** coefficient of variation
**GMI:** glucose management indicator
**$HbA_{1c}$:** glycated hemoglobin
**HBGI:** high blood glucose index
**LBGI:** low blood glucose index
**MAR:** missing at random
**MCAR:** missing completely at random
**MEC-U:** Medical Research Ethics Committees United
**MedAE:** median absolute error
**MNAR:** missing not at random
**RI:** risk index
**TAR:** time above range
**TAR2:** time above range level 2
**TBR:** time below range
**TBR2:** time below range level 2
**TIR:** time in range
**ZGT:** Ziekenhuisgroep Twente

*Edited by T de Azevedo Cardoso; submitted 14.07.23; peer-reviewed by W Jia, T Vandenbroek, A Degraaf; comments to author 19.01.24; revised version received 21.02.24; accepted 10.04.24; published 31.07.24.*

*Please cite as:*
*den Braber N, Braem CIR, Vollenbroek-Hutten MMR, Hermens HJ, Urgert T, Yavuz US, Veltink PH, Laverman GD*
*Consequences of Data Loss on Clinical Decision-Making in Continuous Glucose Monitoring: Retrospective Cohort Study*
*Interact J Med Res 2024;13:e50849*
*URL: https://www.i-jmr.org/2024/1/e50849*
*doi:10.2196/50849*
*PMID:*

Original Paper

# Managing Type 2 Diabetes During the COVID-19 Pandemic: Scoping Review and Qualitative Study Using Systematic Literature Review and Reddit

Meghan S Nagpal[1,2], PEng, MSc; Niloofar Jalali[3], PhD; Diana Sherifali[4], RN, CDE, PhD; Plinio P Morita[1,2,3], PEng, PhD; Joseph A Cafazzo[1,2,5,6], PEng, PhD

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Centre for Digital Therapeutics, Techna Institute, University Health Network, Toronto, ON, Canada

[3]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

[4]School of Nursing, McMaster University, Hamilton, ON, Canada

[5]Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada

[6]Department of Computer Science, University of Toronto, Toronto, ON, Canada

**Corresponding Author:**
Meghan S Nagpal, PEng, MSc
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
Health Sciences Building, 4th Floor
155 College Street
Toronto, ON, M5T 3M6
Canada
Phone: 1 4169784326
Email: meghan.nagpal@gmail.com

## Abstract

**Background:** The COVID-19 pandemic impacted how people accessed health services and likely how they managed chronic conditions such as type 2 diabetes (T2D). Social media forums present a source of qualitative data to understand how adaptation might have occurred from the perspective of the patient.

**Objective:** Our objective is to understand how the care-seeking behaviors and attitudes of people living with T2D were impacted during the early part of the pandemic by conducting a scoping literature review. A secondary objective is to compare the findings of the scoping review to those presented on a popular social media platform Reddit.

**Methods:** A scoping review was conducted in 2021. Inclusion criteria were population with T2D, studies are patient-centered, and study objectives are centered around health behaviors, disease management, or mental health outcomes during the COVID-19 pandemic. Exclusion criteria were populations with other noncommunicable diseases, examining COVID-19 as a comorbidity to T2D, clinical treatments for COVID-19 among people living with T2D, genetic expressions of COVID-19 among people living with T2D, gray literature, or studies not published in English. Bias was mitigated by reviewing uncertainties with other authors. Data extracted from the studies were classified into thematic categories. These categories reflect the findings of this study as per our objective. Data from the Reddit forums related to T2D from March 2020 to early March 2021 were downloaded, and support vector machines were used to classify if a post was published in the context of the pandemic. Latent Dirichlet allocation topic modeling was performed to gather topics of discussion specific to the COVID-19 pandemic.

**Results:** A total of 26 studies conducted between February and September 2020, consisting of 13,673 participants, were included in this scoping literature review. The studies were qualitative and relied mostly on qualitative data from surveys or questionnaires. Themes found from the literature review were “poorer glycemic control,” “increased consumption of unhealthy foods,” “decreased physical activity,” “inability to access medical appointments,” and “increased stress and anxiety.” Findings from latent Dirichlet allocation topic modeling of Reddit forums were “Coping With Poor Mental Health,” “Accessing Doctor & Medications and Controlling Blood Glucose,” “Changing Food Habits During Pandemic,” “Impact of Stress on Blood Glucose Levels,” “Changing Status of Employment & Insurance,” and “Risk of COVID Complications.”

**Conclusions:** Topics of discussion gauged from the Reddit forums provide a holistic perspective of the impact of the pandemic on people living with T2D, which were found to be comparable to the findings of the literature review. The study was limited by only having 1 reviewer for the literature review, but biases were mitigated by consulting authors when there were uncertainties. Qualitative analysis of Reddit forms can supplement traditional qualitative studies of the behaviors of people living with T2D.

***(Interact J Med Res 2024;13:e49073)*** doi:10.2196/49073

**KEYWORDS**

type 2 diabetes; social media; patient-generated health data; big data; machine learning; natural language processing; COVID-19, COVID-19 stress syndrome, health behaviors; Reddit; qualitative; analysis; diabetes; scoping review

## Introduction

### Background

Type 2 diabetes (T2D) is characterized by the body's resistance or insufficient production of insulin. Research suggests that the risks of further complications for people living with T2D can be mitigated through proper self-management [1]. Treatment protocol for proper management of T2D includes glycemic control, weight management, adequate nutrition, regular physical activity, reducing sedentary behaviors, and taking prescribed medications [2].

### COVID-19 and Managing T2D

With the emergence of the COVID-19 pandemic, beginning in March 2020, social distancing measures included business closures, remote school and work measures, prohibition of large crowds, limited socialization outside the household, and increased reliance on digital health care delivery. As a result of these changes and fear of the unknown, stress and anxiety were resulting manifestations [3]. People living with diabetes are already at increased risk for serious complications from COVID-19 due to already being immunocompromised and because the virus may thrive in an environment of high blood glucose [4]. A scoping review conducted in 2023 by Li et al [5] revealed that diabetes prevalence increased among those with severe COVID-19, accounting for 16.8% of deaths. Therefore, it was vital that those living with T2D took extra precautions to avoid the virus. However, proper management of T2D requires healthy lifestyle behaviors, which were likely impacted by the lifestyle changes that occurred during lockdowns, in addition to the exacerbated risk of attaining severe COVID-19 symptoms.

### Study Objective and Rationale

Considering that proper management of T2D requires healthy behaviors and that the implications of the COVID-19 pandemic were disruptive in people's daily lives worldwide, this study aimed to consolidate the literature of studies that examined the health behaviors and attitudes of people living T2D during the first year of the COVID-19 pandemic and to compare the themes gauged from the scoping review to topics of discussion on Reddit forums among people living with T2D during the same period. Social media is a form of patient-generated health data where users can discuss with their peers how they manage T2D through sharing diet, food, symptoms, research, and recipes while obtaining peer support [6]. It also presents a public data source to gauge sentiment and topics of discussion during the initial lockdown period. Our objective was to examine if data from social media, in this case Reddit, provided insights that were similar to findings from the literature review.

## Methods

### Scoping Review

A scoping review was conducted following the framework of Arksey and O'Malley [7] using the following steps: (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) charting the data, and (5) summarizing and reporting the results [8].

### Search Strategy

Searches were conducted in 3 databases (PubMed, Scopus, and CINAHL) from January 2020 to May 2021 (Textbox 1). Using the keywords identified, relevant studies were identified using the inclusion and exclusion criteria from title and abstract to full-text screening. For studies included in the data charting phase, reference lists were scanned for any additional relevant studies. However, these searches did not produce any additional results.

**Textbox 1.** Search terms for the scoping review.

| |
|---|
| *Diabetes* AND (*manag** OR *behave** OR *mental* OR *stress* OR *anxiety* OR *depression*) AND (*COVID* OR *coronavirus* OR *pandemic*) |

### Study Selection

Following the Arksey and O'Malley [7] framework, papers were reviewed in 3 iterations. In the first iteration, abstracts were scanned and selected using the eligibility criteria below. In the second iteration, the full text was scanned using the same eligibility criteria to select papers. Finally, in the third iteration, data were extracted and charted, and studies were excluded if they did not meet the eligibility criteria. Eligibility was determined based on the criteria below and for a paper to be included, all inclusion criteria needed to be met, and not have met any exclusion criteria. Only 1 reviewer (MSN) performed the initial study screening and assessment, but uncertainties about inclusion criteria were addressed to the other authors and

solved through discussion to make the final decision for study eligibility.

The inclusion criteria are (1) the population of focus must include people living with T2D, (2) the findings of the study are patient-centered, and (3) the objectives of the study are to gauge changes in health behaviors, disease management, or mental health outcomes during the COVID-19 pandemic.

The exclusion criteria are (1) people living with other noncommunicable diseases, not as a comorbidity with T2D (an exception to this criterion was made if the population consisted of people with type 1 diabetes or gestational diabetes), (2) the study examines COVID-19 as a comorbidity to T2D, (3) clinical treatments or delivery of care for COVID-19 among people living with T2D, (4) genetic expressions of COVID-19 among people living with T2D, (5) case studies, commentary, review papers, or gray literature (ie, letters to editor, editorials, blogs, and newspapers), or (6) studies not published in English.

## Charting and Extracting Data

To guide data extraction, parameters were created that included the country of study, the time in which the study was conducted, the study sample size, and the main findings. Findings were directly extracted and quoted from the paper and the remaining data parameters were interpreted through analysis from examining the paper. Given the limitations of the study, only 1 reviewer (MSN) was able to conduct the extraction, but clarity was taken from other authors when there were areas of uncertainty.

## Synthesis of Data

The extracted results from the study were examined and were given numerical codes for thematic analysis. Thematic analysis was conducted by the primary reviewer (MSN) and other authors were consulted in the case of uncertainty or discrepancies. Themes were categorized to summarize the studies by their main findings to answer the research question.

## Examination of Reddit Data

### *Data Collection*

For this study, 3 communities on Reddit were examined: r/type2diabetes, r/diabetes_t2, and r/diabetes [9-11]. From the r/diabetes [11] community, only posts that were tagged with the “flair” and “type 2 diabetes” were examined. The former 2 communities are exclusively for people living with T2D, while the latter was only examined if it was tagged as T2D. While there is no way to guarantee that patients living with type 1 diabetes were excluded from this data set, it was reasonable to assume that the discussions in our data set only pertained to T2D, given that they were posted or tagged in communities for people living with T2D.

Reddit was the chosen data source because there were readily available open-source application programming interfaces (APIs) to harness the data through Python scripts. Additionally, because Reddit communities are divided by different interest groups, such as diabetes, it ensured that the data source mostly consisted of the population of interest. Finally, Reddit’s terms and conditions did not forbid the use of data for research purposes and was chosen for that reason [12].

### *Classification of Posts*

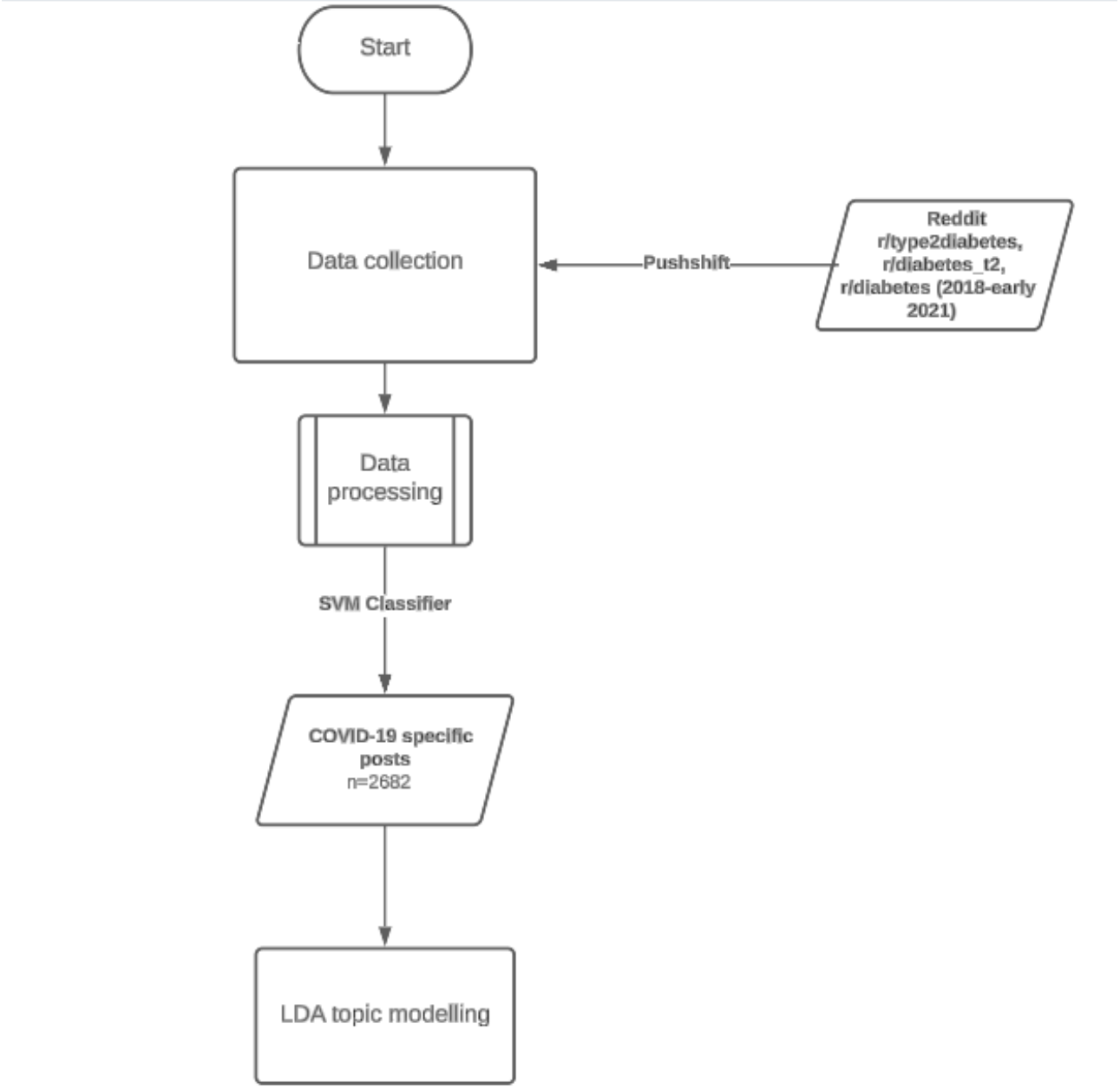
A year’s worth of data were examined, a total of 48,988 posts from March 2020 to March 2021. Within the data set, terms related to the COVID-19 pandemic were manually searched for using search features on Excel (Microsoft Corp) by MSN. These terms included *COVID*, *coronavirus*, *pandemic*, *social distancing*, *lockdown*, *quarantine*, *toilet paper*, *unemploy*, *unemployed*, *work from home*, *working from home*, *telehealth*, *vaccine*, *sanitizer*, and *mask*.

Posts that contained those terms in the text body were manually evaluated for context and labeled as “covid” or “noncovid.” In total, 9803 posts were manually classified by MSN and verified by NJ, with 2065 labeled as “covid” and 7738 labeled as “noncovid” and subsequently classified with the support vector machines. An additional 818 posts that were published in the context of the COVID-19 pandemic were identified, bringing the total number of pandemic-specific posts to 2883. The remainder of unlabeled posts published in the identified pandemic period were labeled as “noncovid.”

### *Data Analysis*

The latent Dirichlet allocation (LDA) topic modeling algorithm [13,14] with the MALLET (Machine Learning for Language Toolkit) package [15] was used to obtain topics of discussion by obtaining clusters of words belonging to a single topic (Figure 1). This unsupervised algorithm was chosen as there was no precedent of topics that were being detected, and thus, there were no data to train a supervised algorithm. After classifying the posts, the entire data set was reprocessed, with 2682 being specific to the COVID-19 pandemic, and topic modeling was performed. A value of *k*, the number of topics, was determined by evaluating the coherence scores outputted by the model for each value of *k* and by manually evaluating the distinction between topics for various *k* values.

**Figure 1.** Collecting and processing of Reddit data. LDA: latent Dirichlet allocation; SVM: support vector machine.




### *Sentiment Analysis*

Sentiment analysis (SA) with the VADER (Valence Aware Dictionary for Sentiment Reasoning) [16] algorithm was performed to understand the subjective emotions, or sentiment, associated with each post. Valence scores of sentiment are calculated on both polarity (if the text is negative or positive) and intensity (how positive or negative a text is), with a normalized compound score, returned between –1 and +1. Thresholds for classifying a text as per VADER are as follows [16]: positive: compound score ≥+0.05; neutral: compound score between –0.05 and +0.05; and negative: compound score ≤–0.05.

## Ethical Considerations

As per the University of Toronto's research ethics guides exemptions in section 1, ethical approval was not obtained for this study because it was assumed that Reddit is a public data source and it is assumed that there is no reasonable expectation of privacy [17]. Moreover, there was no direct interaction between the researcher and the participants, and hence, the researchers did not believe that ethical approval was necessary, as per section 2 [17]. The data used for our qualitative analysis was scraped directly from Reddit and it was assumed to be a public data source. As users do not need a form of authentication to view Reddit forums, it was assumed that the users who posted to them did so with the knowledge that they would be displayed publicly. Data were obtained through a data dump from the Pushshift API, and more than 100 published research studies have already harnessed Reddit data with Pushshift [18]. We do not believe that using this API was a violation of Reddit's API terms of use [19]. However, in a systematic analysis of 727 manuscripts that used Reddit as a data source, only 15% mentioned any form of ethical review [20]. We do acknowledge that there is debate on the ethics of using Reddit data for academic research purposes but nowhere on Reddit's terms and

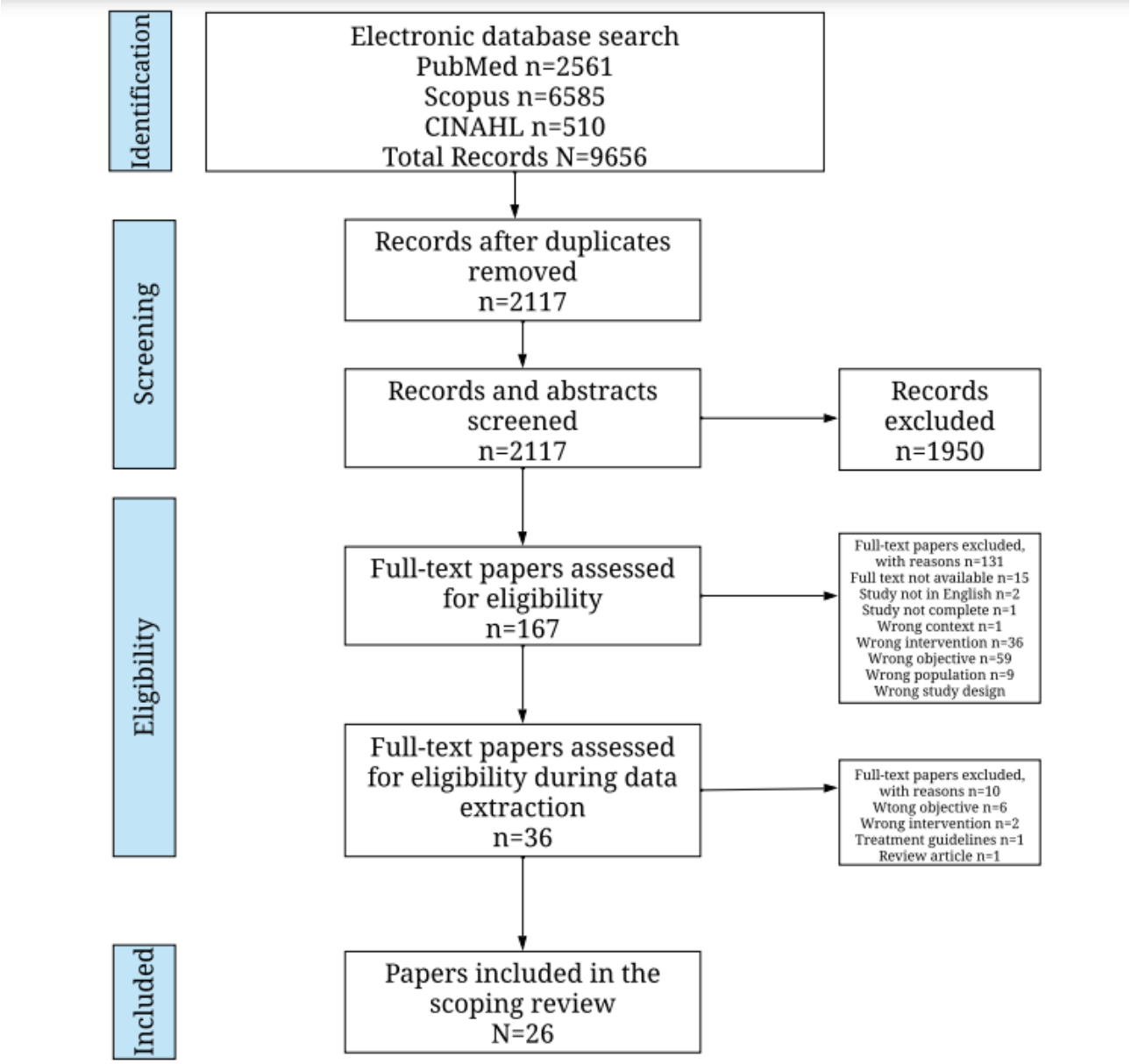
conditions prohibit the use of data for research purposes [12]. Hence, as of the time of research, by omission of information about using data for research purposes, it was assumed that it was within Reddit's terms and conditions. While the username of the post authors was obtained, they were assumed to be pseudonyms of the user and not their actual names. However, we do acknowledge that some users may have integrated their real names into their usernames, but in this study, the usernames were not analyzed to confirm so. Reddit also does not provide identifiable characteristics of individual users, such as their name, gender, or geographical location. However, we do acknowledge that some users may put identifiable information within their text. For this study, this identifiable information was not harnessed or analyzed.

## Results

### Findings From Scoping Review

A total of 9656 papers were identified from the 3 databases. Of these 9656 papers, 7539 (78.1%) papers were duplicates and were subsequently removed. The abstracts of 2117 papers were screened and the full texts of 167 (7.9%) of those papers were scanned to evaluate if they met the inclusion criteria. Of those papers, 36 (1.7%) papers were included for data extraction. Finally, after close examination from data extraction, of the 36 papers, 26 (72.2%) papers were included as part of this scoping review as they met the inclusion criteria. Figure 2 summarizes the process.

**Figure 2.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram from the scoping review.




### Study Characteristics

A total of 26 studies published as of May 2021 were included in this review. With respect to data collection, of the 26 studies, 21 (80%) studies used surveys or questionnaires, 3 (11%) studies used interviews, 2 (8%) studies analyzed blood glucose readings, and 1 (4%) study used hospital consultations. The studies took place between February and September 2020 and included a

sample size of 13,673 people living with diabetes. The geographical representation includes 7 (26%) studies from India, 3 (11%) studies from Japan, 2 (8%) studies from China, 2 (8%) studies from Denmark, 2 (8%) studies from Brazil, 2 (8%) studies from the Netherlands, 1 (4%) study from Turkey, 1 (4%) study from Spain, 1 (4%) study from Germany, 1 (4%) study from Arab Gulf, 1 (4%) study from Mexico, 1 (4%) study from Singapore, 1 (4%) study from Pakistan, and 1 (4%) study from the United Kingdom.

### Findings From Thematic Analysis

The main themes found from the literature review include poor glycemic control, increased consumption of unhealthy foods, reduction in physical activity, and inability to access medical appointments. A detailed summary of all included papers is included in Multimedia Appendix 1.

### Findings From Analysis of Reddit Data

The LDA topic modeling algorithm was performed on 2883 posts published between February 28, 2020, and February 28, 2021.

### Findings From LDA Topic Modeling

Table 1 summarizes the topics and associated words for COVID-19–related posts.

**Table 1.** Topics and associated words for COVID-19–related posts.

| Topic number | Topic name | Words |
|---|---|---|
| Topic 1 | Coping With Poor Mental Health | feel, day, week, work, weight, lose, walk, bit, symptom, bad, lot, gym, night, happen, end, great, covid, felt, ill, diagnosis |
| Topic 2 | Accessing Doctor & Medications and Controlling Blood Glucose | doctor, test, ac, year, insulin, month, metformin, diagnose, low, glucose, diet, week, med, exercise, change, level, stop, reading, check, medication |
| Topic 3 | Changing Food Habits During Pandemic | eat, food, carbs, lot, meal, low, carb, diet, hard, hour, thing, water, keto, fast, cut, easy, glucose, add, stuff, rice |
| Topic 4 | Impact of Stress on Blood Glucose Levels | blood, glucose, high, time, stress, number, long, exercise, bg, body, problem, morning, normal, control, sleep, make, level, effect, change, kind |
| Topic 5 | Changing Status of Employment & Insurance | work, home, hospital, today, time, year, talk, give, advice, wait, insurance, visit, guess, state, order, meter, live, strip, situation, friend |
| Topic 6 | Risk of COVID Complications | covid, type, health, care, risk, sick, control, issue, virus, disease, diabetic, question, pandemic, mask, hand, case, infection, wear, patient, home |

### Findings From Sentiment Analysis

Impact of Stress on Blood Glucose Levels and Coping With Poor Mental Health had average compound scores that fell in the threshold of being classified as neutral (odds ratio [OR] 0.0252, 95% CI –0.0344 to 0.0849 and OR 0.0492, 95% CI –0.0121 to 0.1105, respectively). Risk of COVID Complications, Accessing Doctor & Medications and Controlling Blood Glucose, and Changing Status of Employment & Insurance were ranked next from lowest to highest with average compound scores of OR 0.0876 (95% CI 0.0322-0.1430), OR 0.1457 (95% CI 0.0879-0.2035), and OR 0.1748 (95% CI 0.1263-0.2232), respectively, and finally, Changing Food Habits During Pandemic had the highest average compound score of 0.2544 (95% CI 0.1965-0.3123).

## Discussion

### Principal Results

Our analysis revealed that people living with T2D were negatively impacted by the pandemic mentally and were negatively impacted by how they managed their chronic disease. Our literature review found that people living with T2D were negatively impacted by the pandemic by having poorer glycemic control, poorer lifestyle behaviors, their inability to access medical appointments, and increased stress and anxiety. Our analysis of Reddit data found similar themes, with additional emphasis on the economic impacts of the pandemic among people living with T2D.

Our literature review found that the COVID-19 pandemic impacted people's lives through poorer glycemic control, increased consumption of unhealthy foods, decreases in physical activity, inability to access medical appointments, and increased stress and anxiety toward the impact of the lockdown and fear of being exposed to the coronavirus. Topic modeling from data on 3 Reddit forums for people living with T2D found the following topics: Coping With Poor Mental Health, Accessing Doctor & Medications and Controlling Blood Glucose, Changing Food Habits During Pandemic, Impact of Stress on Blood Glucose Levels, Changing Status of Employment & Insurance, and Risk of COVID Complications. The additional finding of employment as a topic of discussion on Reddit forums suggests that digital discussion presents a holistic perspective of diabetes management that considers the person's life as a whole when managing their disease.

Furthermore, the majority of the 26 studies included in our literature review mostly relied on surveys and interviews to obtain their data. Surveys and interviews are often time-consuming processes. However, analyzing data from forums such as Reddit, using machine learning algorithms such as topic modeling and SA, can be a quicker method to obtain a broad range of themes and sentiments from a large volume of participants when performing qualitative research. We do not suggest that qualitative analysis from digital forums could replace traditional qualitative research, but can rather supplement it as our study demonstrates that the results from

our analysis are comparable to results from traditional qualitative studies of people living with T2D.

### Glycemic Control

As glycemic control is a major component of the self-management of T2D, it was expected that this would be a major theme found both in our literature review and in our analysis of Reddit data. While psychological stress is subjective among individuals, few studies have demonstrated that psychological stressors have been linked to hyperglycemia [21-23]. Considering that the major theme of our findings suggested that the pandemic was a stressor for people living with T2D, it was hypothesized that this stress would have an impact on glycemic control. From our Reddit analysis, the topic Impact of Stress on Blood Glucose Levels had the lowest sentiment score of OR 0.0252 (95% CI –0.0344 to 0.0849), suggesting that there was increased anxiety toward managing blood glucose levels.

Our literature review further reiterated that people living with T2D experienced increased blood glucose or increased hemoglobin $A_{1c}$ ($HbA_{1c}$) levels during the pandemic [24-28]. While 1 study did correlate higher $HbA_{1c}$ with increased levels of stress during the pandemic [29], our literature review also attributed poorer glycemic control as a result of reduced blood glucose monitoring and reduced medical visits being reasons for this [24,30-35]. This sentiment was also reflected in our analysis of Reddit data. The topic Accessing Doctor & Medications and Controlling Blood Glucose considered that reduced access to medication and health care providers impacts blood glucose levels, and while reduced health care visits were a factor in reduced glycemic control, the Reddit analysis additionally considered that losing employment was also a stressor for people living with T2D and also resulted in lost insurance benefits that reduced doctor visits and medication access.

### Lifestyle Management

As T2D is managed by lifestyle behaviors, they were nonetheless a significant theme of discussion found in both our literature review and our analysis of T2D Reddit communities. Our literature review revealed that lifestyle behaviors among people living with T2D were impacted by increased consumption of unhealthy foods and reduced physical activity [24-27,30,33,36-38]. As a result of these behavior changes, participants reported changes in body weight [27,37].

These findings were supported by our analysis of Reddit as Changing Food Habits During Pandemic was a topic of discussion. Interestingly, this topic was associated with the highest average sentiment (OR 0.2544, 95% CI 0.1965-0.3123). While the literature suggests that dietary changes were attributed to stress during the pandemic [33], the positive sentiment score of the Reddit posts may be an indication that increased unhealthy food consumption may have been a coping mechanism associated with positive emotions through the stressful time and that there was a sense of camaraderie and bonding among peers through this coping mechanism.

Considering physical activity, in the topic Coping With Poor Mental Health, the terms “gym” and “walk” were included. This may suggest that stress was related to gym closures in the initial months of the pandemic and people relying on walking as a means of physical activity. The literature review supports that gym closures and having fewer opportunities to walk due to teleworking and closures of businesses are attributed to reducing physical activity [25,27]. Moreover, the literature review attributes changes in exercise behaviors as a result of pandemic-related stress [38], and hence, it was fitting that the topic modeling algorithm pooled terms related to mental health with terms related to physical activity.

### Access to Diabetes Care

During the COVID-19 pandemic, accessing care was perceived as a significant barrier to managing T2D as more clinical visits were done through telehealth as a means to protect patients and health care providers from exposure to the coronavirus. However, not all people living with T2D had access to telehealth care, particularly those living in rural communities [25]. Moreover, health care providers were called to aid in treating patients infected with COVID-19 [25], resulting in people living with T2D being unable to see their regular health care provider or being treated by health care providers who were not experienced in managing T2D [39]. Our literature review revealed that people living with T2D had difficulty managing their blood glucose levels and felt depression as a result of missed medical appointments.

Among our analysis of Reddit data, Access to Doctor & Controlling Blood Glucose was identified as a topic of discussion among people living with T2D, with an average sentiment score (OR) of 0.1457 (95% CI 0.0879-0.2035). Further examination of Reddit data revealed that while barriers to accessing health care providers existed, another barrier was presented through the fear of acquiring COVID-19 infection and avoiding hospitalization in potentially dangerous situations. The fear of acquiring COVID-19 infection was also reflected in our literature review in a general sense. Further research in the years ahead would need to examine the impact of the lessening of in-person health care visits among people living with T2D.

### Mental Health

With the COVID-19 pandemic being disruptive to personal lives worldwide, many people experienced elevated stress and anxiety. The mental health impact of this pandemic is expected to be long-term due to the extreme measures that were necessary to prevent the spread of the virus and the resulting economic implications [40]. Our literature review revealed that people living with T2D were no exception to the stressors of the pandemic which included social isolation [28,29,41] and financial stress [28,29]. However, stressors that were specific to people living with T2D included missing medical appointments [42], being unable to access medications and supplies to manage diabetes [42,43], and managing their disease [29]. Additionally, people living with T2D were anxious about being exposed to COVID-19 as they felt that they were more vulnerable to serious complications or death [26,31,41-44].

Comparing these findings from our literature review to our analysis of Reddit data found Coping With Poor Mental Health,

Impact of Stress on Blood Glucose Levels, and Risk of COVID Complications as topics of discussion. Overall, these topics were associated with lower sentiment scores (OR 0.0492, 95% CI –0.0121 to 0.1105; OR 0.0252, 95% CI –0.0344 to 0.0849; and OR 0.0876, 95% CI 0.0322 to 0.1430; respectively). Specifically, within the Reddit analysis, users acknowledge that increased stress during the pandemic impacted their glycemic control but users also felt that they would be helpless if they experienced diabetes-related complications as hospitalization would put them at risk of acquiring a COVID-19 infection. Moreover, as supported in the literature review, people living with T2D were generally afraid of being exposed to the virus with their increased risk state. While our study demonstrated that people living with T2D were using peer support as a means to cope with the stressors of the pandemic, it also demonstrates that users were negatively impacted by the psychological stressors of the pandemic.

### Impact to Employment

Efforts to curb the virus resulted in many employers worldwide requiring their employees to work from home [45], which resulted in changes in work-life balance and mental health issues with the inability to interact with others outside the household [46,47]. Moreover, as businesses shut down during the pandemic to curb the spread of the virus, the pandemic resulted in the loss of employment for many workers. The unemployment rate reached 14.1% in the United States in April 2020, the highest since data collection began in 1948 [48]. Additionally, only 60.2% of the labor force participated in April 2020, the lowest participation observed since the 1970s [48]. Unemployment already poses the issue of loss of income and standard of living and decreased sense of self-purpose [49], potentially impacting health behaviors due to increased stress. However, many people unemployed were also impacted by changes to their health insurance because of their job loss [50,51].

While Kishimoto et al [25] suggested that teleworking resulted in reduced physical activity among people living with T2D, there was little discussion about the financial and employment implications of the pandemic among people living with T2D. However, Changing Status of Employment & Insurance was a topic of discussion among Reddit users on T2D forums. From the Reddit discussions, it could be inferred that a loss of insurance posed a barrier for people living with T2D to get medication and blood glucose meters and strips, affecting glycemic control. Clinicians must consider how one's employment status affects people living with T2D, especially when they are additionally posed with a barrier to accessing clinical care, and outline treatment options in these situations.

### Comparison to Prior Work

This study builds upon previous studies that have harnessed data from weblogs and social media websites to understand diabetes behaviors and the sentiment associated with the texts. People living with T2D use social media and digital forums to discuss their condition and related information among their peers [52]. These themes of discussion include diet, food, symptoms, research, recipes, and news [52-55]. This study uncovered similar themes of discussion through analysis of Reddit forums, with the added context of the COVID-19 pandemic, given that this study was conducted in 2021, as the pandemic was ravaging globally. However, no other study at the time compared the findings of social media analysis to the findings from traditional qualitative studies. Our study suggests that social media can be a supplemental data source when performing clinical qualitative analysis.

### Comparison of Literature Review to Reddit Analysis

Our literature review summarized 26 studies conducted on 13,673 people living with diabetes to understand how the pandemic impacted and how they managed their diseases. Our analysis of Reddit data used support vector machines to classify Reddit posts written from 2020 to early 2021, published in the context of the pandemic from 1263 distinct authors. Most of the data obtained by researchers of the studies included in the literature review were through surveys and interviews, while the data obtained in the Reddit analysis used APIs to scrape data that were posted by Reddit users and analyzed using LDA topic modeling and VADER SA. As displayed in Textbox 2, the topics gauged from both studies were comparable to one another, with the Reddit analysis gauging an additional topic of Changing Status of Employment & Insurance. This additional finding suggests that discussions on Reddit offer insight from a holistic perspective that considers aspects of a person's life in the context of their disease, beyond treating symptoms. Moreover, analyses from Reddit forums can be less time-consuming than conducting long surveys or interviews and can collect data from a larger volume of users than from a single study.

**Textbox 2.** Comparison of topics found in the literature review to topics found in the Reddit discussion.

**Literature review topics**

- Increased Consumption of Unhealthy Foods
- Decreased Physical Activity
- Inability to Access Medical Appointments
- Anxiety Toward Lockdown
- Fear of COVID Exposure

**Reddit topics**

- Changing Food Habits During Pandemic
- Impact of Stress on Blood Glucose Levels
- Accessing Doctor & Medications and Controlling Blood Glucose
- Coping With Poor Mental Health
- Risk of COVID Complications
- Changing Status of Employment & Insurance

## Limitations

Due to the staffing limitations and reliability of remote work during the pandemic, only 1 reviewer, MSN, was able to conduct the study search and study data extraction. However, any uncertainties were addressed to DS and JAC to mitigate any risk of bias.

Topic modeling and SA were performed on the Reddit posts but the posts were not thoroughly examined for context, and hence, the authors cannot comment on the quality of the discussions posted on the forum. We cannot confirm that the users who published the posts were all people living with T2D. We only examined the 3 Reddit communities mentioned in the study and no other subreddits about the coronavirus or mental health. Furthermore, we had no information about the demographics of the users of the diabetes forums of Reddit and assumed that the demographics were similar to the demographics of all Reddit users. Under this assumption, there could be a sampling bias in our Reddit data as Reddit users are mostly male and fall into the age demographic of 18-49 years [56,57]. Geographically, the United States has the largest number of Reddit users, while other users mostly reside in English-speaking, higher-income countries [58].

## Conclusions

The findings from the literature review included topics of glycemic control, lifestyle management, access to diabetes care, and the impact on mental health among people living with T2D. However, an examination of Reddit data revealed an additional theme of employment being impacted during the pandemic, affecting diabetes lifestyle behaviors. Moreover, Reddit presented a large sample size of participants. Therefore, social media presents an opportunity to holistically observe the behaviors of those managing chronic diseases.

## Acknowledgments

This study was conducted as part of MSN's master's thesis project and funding was provided by the Centre for Digital Therapeutics. No generative artificial intelligence was used for this paper.

## Data Availability

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

## Authors' Contributions

This study was conducted by MSN. MSN received guidance on the methodology from NJ and PPM. DS served as the clinical advisor for this study. All authors contributed to conceptualization. Data curation was conducted by MSN, with assistance from NJ. The methodology was designed by PPM. Formal analysis was conducted by MSN, with assistance from NJ and PPM. The investigation was conducted by MSN. Validation and verification were provided by DS. Funding support, project administration, and supervision were provided by JAC. This paper was originally written and edited by MSN. All listed authors approved the final version of this paper.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Findings from scoping review.
[DOCX File , 27 KB - ijmr_v13i1e49073_app1.docx ]

Multimedia Appendix 2
PRISMA checklist.
[DOCX File , 85 KB - ijmr_v13i1e49073_app2.docx ]

## References


1. Sherifali D, Viscardi V, Bai J, Ali RMU. Evaluating the effect of a diabetes health coach in individuals with type 2 diabetes. Can J Diabetes 2016;40(1):84-94. [doi: 10.1016/j.jcjd.2015.10.006] [Medline: 26827684]
2. Diabetes Canada Clinical Practice Guidelines Expert Committee. Diabetes Canada 2018 clinical practice guidelines for the prevention and management of diabetes in Canada. Can J Diabetes 2018;42(Suppl 1):A1-A18-S1-S326. [doi: 10.1016/s1499-2671(17)31026-2]
3. Mental health and COVID-19. WHO. URL: https://www.who.int/teams/mental-health-and-substance-use/mental-health-and-covid-19 [accessed 2022-01-24]
4. International Diabetes Federation—Home. COVID-19 and diabetes. URL: https://www.idf.org/aboutdiabetes/what-is-diabetes/covid-19-and-diabetes/1-covid-19-and-diabetes.html [accessed 2022-01-24]
5. Li R, Shen M, Yang Q, Fairley CK, Chai Z, McIntyre R, et al. Global diabetes prevalence in COVID-19 patients and contribution to COVID-19-related severity and mortality: a systematic review and meta-analysis. Diabetes Care 2023;46(4):890-897 [FREE Full text] [doi: 10.2337/dc22-1943] [Medline: 36826982]
6. Nagpal MS, Barbaric A, Sherifali D, Morita PP, Cafazzo JA. Patient-generated data analytics of health behaviors of people living with type 2 diabetes: scoping review. JMIR Diabetes 2021;6(4):e29027 [FREE Full text] [doi: 10.2196/29027] [Medline: 34783668]
7. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
8. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
9. r/type2diabetes. Reddit. URL: https://www.reddit.com/r/type2diabetes/ [accessed 2021-08-24]
10. r/diabetes_t2. Reddit. URL: https://www.reddit.com/r/diabetes_t2/ [accessed 2021-08-24]
11. r/diabetes. Reddit. URL: https://www.reddit.com/r/diabetes/ [accessed 2021-08-24]
12. Reddit User Agreement. URL: https://www.redditinc.com/policies/user-agreement-september-12-2021 [accessed 2021-09-23]
13. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. JMLR 2003:993-1022. [doi: 10.7551/mitpress/1120.003.0082]
14. Blei DM. Probabilistic topic models. Commun ACM 2012;55(4):77-84. [doi: 10.1145/2133806.2133826]
15. Topic modeling. URL: http://mallet.cs.umass.edu/topics.php [accessed 2021-08-24]
16. vaderSentiment. PyPI. URL: https://pypi.org/project/vaderSentiment/ [accessed 2021-08-24]
17. Activities exempt from human ethics review. University of Toronto. URL: https://research.utoronto.ca/ethics-human-research/activities-exempt-human-ethics-review [accessed 2024-05-30]
18. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The Pushshift Reddit dataset. 2020 Presented at: Proceedings of the International AAAI Conference on Web and Social Media; June 3–6, 2024; Buffalo, New York, USA p. 830-839 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7347
19. API terms of use. Reddit. 2016. URL: https://www.reddit.com/wiki/api-terms [accessed 2021-09-29]
20. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying reddit: a systematic overview of disciplines, approaches, methods, and ethics. Social Media + Society 2021;7(2):205630512110190. [doi: 10.1177/20563051211019004]
21. Goetsch VL, VanDorsten B, Pbert LA, Ullrich IH, Yeater RA. Acute effects of laboratory stress on blood glucose in noninsulin-dependent diabetes. Psychosom Med 1993;55(6):492-496. [doi: 10.1097/00006842-199311000-00004] [Medline: 8310109]
22. Faulenbach M, Uthoff H, Schwegler K, Spinas GA, Schmid C, Wiesli P. Effect of psychological stress on glucose control in patients with type 2 diabetes. Diabet Med 2012;29(1):128-131. [doi: 10.1111/j.1464-5491.2011.03431.x] [Medline: 21883440]
23. Bruce DG, Chisholm DJ, Storlien LH, Kraegen EW, Smythe GA. The effects of sympathetic nervous system activation and psychological stress on glucose metabolism and blood pressure in subjects with type 2 (non-insulin-dependent) diabetes mellitus. Diabetologia 1992;35(9):835-843. [doi: 10.1007/BF00399929] [Medline: 1397778]
24. Khader MA, Jabeen T, Namoju R. A cross sectional study reveals severe disruption in glycemic control in people with diabetes during and after lockdown in India. Diabetes Metab Syndr 2020;14(6):1579-1584 [FREE Full text] [doi: 10.1016/j.dsx.2020.08.011] [Medline: 32858476]
25. Kishimoto M, Ishikawa T, Odawara M. Behavioral changes in patients with diabetes during the COVID-19 pandemic. Diabetol Int 2021;12(2):241-245 [FREE Full text] [doi: 10.1007/s13340-020-00467-1] [Medline: 33020726]

26. Sankar P, Ahmed WN, Koshy VM, Jacob R, Sasidharan S. Effects of COVID-19 lockdown on type 2 diabetes, lifestyle and psychosocial health: a hospital-based cross-sectional survey from South India. Diabetes Metab Syndr 2020;14(6):1815-1819 [FREE Full text] [doi: 10.1016/j.dsx.2020.09.005] [Medline: 32956926]
27. Tanaka N, Hamamoto Y, Kurotobi Y, Yamasaki Y, Nakatani S, Matsubara M, et al. Lifestyle changes as a result of COVID-19 containment measures: bodyweight and glycemic control in patients with diabetes in the Japanese declaration of a state of emergency. J Diabetes Investig 2021;12(9):1718-1722 [FREE Full text] [doi: 10.1111/jdi.13526] [Medline: 33599073]
28. Khare J, Jindal S. Observational study on effect of lock down due to COVID 19 on HBA1c levels in patients with diabetes: experience from central India. Prim Care Diabetes 2022;16(6):775-779 [FREE Full text] [doi: 10.1016/j.pcd.2020.12.003] [Medline: 33419712]
29. Fisher L, Polonsky W, Asuni A, Jolly Y, Hessler D. The early impact of the COVID-19 pandemic on adults with type 1 or type 2 diabetes: a national cohort study. J Diabetes Complications 2020;34(12):107748 [FREE Full text] [doi: 10.1016/j.jdiacomp.2020.107748] [Medline: 33059981]
30. Ghosh A, Arora B, Gupta R, Anoop S, Misra A. Effects of nationwide lockdown during COVID-19 epidemic on lifestyle and other medical issues of patients with type 2 diabetes in North India. Diabetes Metab Syndr 2020;14(5):917-920 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.044] [Medline: 32574982]
31. Nachimuthu S, Vijayalakshmi R, Sudha M, Viswanathan V. Coping with diabetes during the COVID-19 lockdown in India: results of an online pilot survey. Diabetes Metab Syndr 2020;14(4):579-582 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.053] [Medline: 32416527]
32. Yeoh E, Tan SG, Lee YS, Tan HH, Low YY, Lim SC, et al. Impact of COVID-19 and partial lockdown on access to care, self-management and psychological well-being among people with diabetes: a cross-sectional study. Int J Clin Pract 2021 Aug;75(8):e14319 [FREE Full text] [doi: 10.1111/ijcp.14319] [Medline: 33974316]
33. Tiwari A, Kumar D, Ansari MS, Chaubey SK, Gupta NR, Agarwal V, et al. Impact of lockdown on self-care management among patients with type 2 diabetes mellitus residing in Lucknow city, India—a cross-sectional study. Clin Epidemiol Glob Health 2021;10:100703. [doi: 10.1016/j.cegh.2021.100703]
34. Silva-Tinoco R, González-Cantú A, de la Torre-Saldaña V, Guzmán-Olvera E, Cuatecontzi-Xochitiotzi T, Castillo-Martínez L, et al. Effect in self-care behavior and difficulties in coping with diabetes during the COVID-19 pandemic. Rev Mex Endocrinol Metab Nutr 2021;8(1):13. [doi: 10.24875/rme.20000063]
35. Bala R, Srivastava A, Potsangbam T, Anal L, Ningthoujam GD. Self care practices and psychological distress among diabetic patients in Manipur during COVID-19: a scenario from the North East. Diabetes Metab Syndr 2021;15(1):93-98 [FREE Full text] [doi: 10.1016/j.dsx.2020.12.015] [Medline: 33316765]
36. Ruiz-Roso MB, Knott-Torcal C, Matilla-Escalante DC, Garcimartín A, Sampedro-Nuñez MA, Dávalos A, et al. COVID-19 lockdown and changes of the dietary pattern and physical activity habits in a cohort of patients with type 2 diabetes mellitus. Nutrients 2020;12(8):2327 [FREE Full text] [doi: 10.3390/nu12082327] [Medline: 32759636]
37. Munekawa C, Hosomi Y, Hashimoto Y, Okamura T, Takahashi F, Kawano R, et al. Effect of coronavirus disease 2019 pandemic on the lifestyle and glycemic control in patients with type 2 diabetes: a cross-section and retrospective cohort study. Endocr J 2021;68(2):201-210 [FREE Full text] [doi: 10.1507/endocrj.EJ20-0426] [Medline: 32999133]
38. Regeer H, Nieuwenhuijse EA, Vos RC, Kiefte-de Jong JC, van Empelen P, de Koning EJP, et al. Psychological factors associated with changes in physical activity in Dutch people with type 2 diabetes under societal lockdown: a cross-sectional study. Endocrinol Diabetes Metab 2021;4(3):e00249 [FREE Full text] [doi: 10.1002/edm2.249] [Medline: 34277973]
39. Shi C, Zhu H, Liu J, Zhou J, Tang W. Barriers to self-management of type 2 diabetes during COVID-19 medical isolation: a qualitative study. Diabetes Metab Syndr Obes 2020;13:3713-3725 [FREE Full text] [doi: 10.2147/DMSO.S268481] [Medline: 33116721]
40. Iqbal SZ, Li B, Onigu-Otito E, Naqvi MF, Shah AA. The long-term mental health effects of COVID-19. Psychiatr Ann 2020;50(12):522-525. [doi: 10.3928/00485713-20201103-01]
41. Madsen KP, Willaing I, Rod NH, Varga TV, Joensen LE. Psychosocial health in people with diabetes during the first three months of the COVID-19 pandemic in Denmark. J Diabetes Complications 2021;35(4):107858 [FREE Full text] [doi: 10.1016/j.jdiacomp.2021.107858] [Medline: 33573890]
42. Al-Sofiani ME, Albunyan S, Alguwaihes AM, Kalyani RR, Golden SH, Alfadda A. Determinants of mental health outcomes among people with and without diabetes during the COVID-19 outbreak in the Arab Gulf region. J Diabetes 2021;13(4):339-352. [doi: 10.1111/1753-0407.13149] [Medline: 33351239]
43. Joensen LE, Madsen KP, Holm L, Nielsen KA, Rod MH, Petersen AA, et al. Diabetes and COVID-19: psychosocial consequences of the COVID-19 pandemic in people with diabetes in Denmark—what characterizes people with high levels of COVID-19-related worries? Diabet Med 2020;37(7):1146-1154 [FREE Full text] [doi: 10.1111/dme.14319] [Medline: 32392380]
44. Basit KA, Zafar AB, Fawwad A, Waris N, Shaheen F, Basit A. Psychometric analysis for fear of COVID-19 scale (FCV-19S) and its association with depression in patients with diabetes: a cross sectional study from a tertiary care centre in Karachi, Pakistan. Diabetes Metab Syndr 2021;15(3):733-737 [FREE Full text] [doi: 10.1016/j.dsx.2021.03.008] [Medline: 33813249]

45. Savić D. COVID-19 and work from home: digital transformation of the workforce. Grey J (TGJ) 2020;16(2):101-104 [FREE Full text]
46. Killgore WDS, Cloonan SA, Taylor EC, Dailey NS. Loneliness: a signature mental health concern in the era of COVID-19. Psychiatry Res 2020;290:113117 [FREE Full text] [doi: 10.1016/j.psychres.2020.113117] [Medline: 32480121]
47. Fana M, Torrejón Pérez S, Fernández-Macías E. Employment impact of COVID-19 crisis: from short term effects to long terms prospects. J Ind Bus Econ 2020;47(3):391-410. [doi: 10.1007/s40812-020-00168-5]
48. Unemployment Rates During the COVID-19 Pandemic. Washington, DC: Congressional Research Service; 2021.
49. Issue briefing: unemployment and mental health. Institute for Work & Health. 2009. URL: https://www.iwh.on.ca/sites/iwh/files/iwh/reports/iwh_issue_briefing_mental_health_2009.pdf [accessed 2024-07-03]
50. Gangopadhyaya A, Garrett AB. Unemployment, health insurance, and the COVID-19 recession. SSRN J 2020:1-8. [doi: 10.2139/ssrn.3568489]
51. Garrett AB, Gangopadhyaya A. How the COVID-19 recession could affect health insurance coverage. SSRN 2020:1-12. [doi: 10.2139/ssrn.3598558]
52. Lincke A, Lundberg J, Thunander M, Milrad M, Lundberg J, Jusufi I. Diabetes information on social media internet. 2018 Presented at: VINCI '18: Proceedings of the 11th International Symposium on Visual Information Communication and Interaction; August 13, 2018; Växjö, Sweden. [doi: 10.1145/3231622.3232508]
53. Abbar S, Mejova Y, Weber I. You tweet what you eat: studying food consumption through twitter. USA: Association for Computing Machinery; 2015 Presented at: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18, 2015; New York, NY. [doi: 10.1145/2702123.270215]
54. Oyebode O, Orji R. Detecting factors responsible for diabetes prevalence in Nigeria using social media and machine learning. 2019 Presented at: 15th International Conference on Network and Service Management (CNSM); October 21-25, 2019; Halifax, NS, Canada. [doi: 10.23919/CNSM46954.2019.9012679]
55. Griffis H, Asch DA, Schwartz HA, Ungar L, Buttenheim AM, Barg FK, et al. Using social media to track geographic variability in language about diabetes: analysis of diabetes-related tweets across the United States. JMIR Diabetes 2020;5(1):e14431 [FREE Full text] [doi: 10.2196/14431] [Medline: 32044757]
56. Dean B. Reddit usage and growth statistics: how many people use Reddit in 2021. Backlinko. URL: https://backlinko.com/reddit-users [accessed 2021-08-24]
57. Agrawal A. The user demographics of Reddit: the official app. Medium. URL: https://medium.com/@sm_app_intel/the-user-demographics-of-reddit-the-official-app-7e2e18b1e0e1 [accessed 2021-08-24]
58. Regional distribution of desktop traffic to Reddit.com as of December 2023, by country. Statista. URL: https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/ [accessed 2021-08-24]


## Abbreviations

**API:** application programming interface
**HbA1c:** hemoglobin A1c
**LDA:** latent Dirichlet allocation
**MALLET:** Machine Learning for Language Toolkit
**OR:** odds ratio
**SA:** sentiment analysis
**T2D:** type 2 diabetes
**VADER:** Valence Aware Dictionary for Sentiment Reasoning


*Edited by T de Azevedo Cardoso; submitted 16.05.23; peer-reviewed by M Sheth; comments to author 10.02.24; revised version received 18.04.24; accepted 07.06.24; published 08.08.24.*

*Please cite as:*
*Nagpal MS, Jalali N, Sherifali D, Morita PP, Cafazzo JA*
*Managing Type 2 Diabetes During the COVID-19 Pandemic: Scoping Review and Qualitative Study Using Systematic Literature Review and Reddit*
*Interact J Med Res 2024;13:e49073*
*URL: https://www.i-jmr.org/2024/1/e49073*
*doi:10.2196/49073*
*PMID:*

Original Paper

# 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 Sallam[1,2,3], MD, PhD; Muna Barakat[4], PhD; Mohammed Sallam[5], MSc, PharmD

[1]Department of Pathology, Microbiology and Forensic Medicine, School of Medicine, The University of Jordan, Amman, Jordan

[2]Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman, Jordan

[3]Department of Translational Medicine, Faculty of Medicine, Lund University, Malmo, Sweden

[4]Department of Clinical Pharmacy and Therapeutics, Faculty of Pharmacy, Applied Science Private University, Amman, Jordan

[5]Department 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

helpful base to establish a universally accepted approach to standardize the design and reporting of generative AI-based studies in health care, which is a swiftly evolving research topic.

***(Interact J Med Res 2024;13:e54704)*** doi:10.2196/54704

**KEYWORDS**

guidelines; evaluation; meaningful analytics; large language models; decision support

## 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 DOCTYPE (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 rxiv 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.

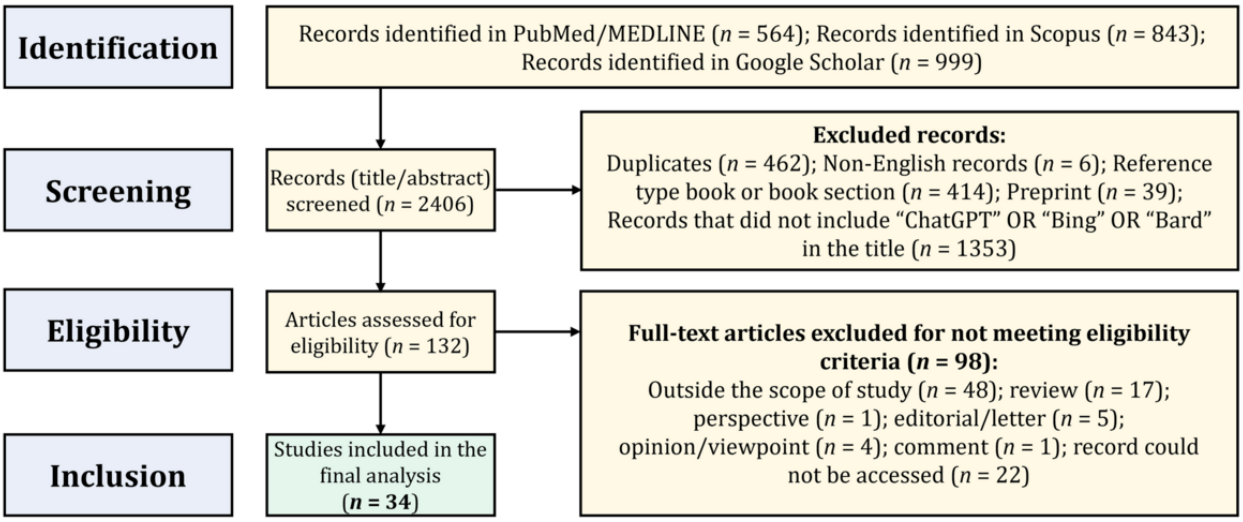
## Results

### 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.

**Figure 1.** The approach for selecting articles.




### 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.

| Authors | Model | Timing | Count | Specificity of the prompts and language |
|---|---|---|---|---|
| Al-Ashwal et al [42] | ChatGPT-3.5, ChatGPT-4, Bing, and Bard with unclear settings | One month, May 2023 | 255 drug-drug pairs | The exact prompts were provided in English. |
| Alfertshofer et al [43] | ChatGPT with unclear settings | Not provided | 1800 questions | 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. |
| Ali et al [44] | ChatGPT Feb 9 free version with unclear settings | Not provided | 50 items | Information was provided fully in the supplementary file of the article in English. |
| Aljindan et al [45] | ChatGPT-4 with unclear settings | Not provided | 220 questions | Initial prompting that involved role playing as a medical professional. The language was English. |
| Altamimi et al [46] | ChatGPT-3.5 | Single day not otherwise specified | 9 questions | The exact prompts were provided in English. |
| Baglivo et al [47] | Bing, ChatGPT, Chatsonic, Bard, and YouChat with full details of the mode and large language model, including plugins | Exact dates were provided for each model (April 12, 13, and 14, 2023, and July 13, 2023) | 15 questions | Italian was used. |
| Biswas et al [48] | ChatGPT-3.5 with unclear settings | Exact date provided (March 16, 2023) | 11 questions | The exact prompts were provided in English, with a new session for each question. |
| Chen et al [49] | ChatGPT-4 with unclear settings | Not provided | 560 questions | The exact prompts were provided in English. |
| Deiana et al [50] | ChatGPT-3.5 and ChatGPT-4 with unclear settings | Not provided | 11 questions | 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. |
| Fuchs et al [51] | ChatGPT-3 and ChatGPT-4 with unclear settings | Exact dates provided (February 19, 2023, and March 25, 2023) | 60 questions | 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. |
| Ghosh & Bir [52] | ChatGPT (version March 14, 2023) with unclear settings | March 14 and 16, 2023 | 200 questions | 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. |
| Giannos [53] | ChatGPT-3 and ChatGPT-4 with unclear settings | Not provided | 69 questions | Not explicitly provided. |
| Gobira et al [54] | ChatGPT-4 with unclear settings | Not provided | 125 questions | Portuguese was used. |
| Grewal et al [55] | ChatGPT-4 with unclear settings | The first week of May 2023 | Not clear | The exact prompts were provided in English. One follow-up prompt was used for enhancement of some prompts. |
| Guerra et al [56] | ChatGPT-4 with unclear settings | Not provided | 591 questions | The exact prompts were provided, while the language was not explicitly provided. |
| Hamed et al [57] | ChatGPT-4 with unclear settings | Not provided | Not clear | The exact prompts and language were not explicitly provided. Different prompts were tried to identify the most suitable. |
| Hoch et al [58] | ChatGPT (May 3rd version) with unclear settings | May 5 and 7, 2023 | 2576 questions | The exact prompts were provided, while the language was not explicitly provided. |

| Authors | Model | Timing | Count | Specificity of the prompts and language |
|---|---|---|---|---|
| Juhi et al [59] | ChatGPT with unclear settings | February 20, 2023, to March 5, 2023 | 40 drug-drug pairs | The exact prompts were provided in English. |
| Kuang et al [60] | ChatGPT with unclear settings | Not provided | Not clear | The exact prompts were not explicitly provided. English was used. |
| Kumari et al [61] | ChatGPT-3.5, Bard, and Bing with unclear settings | July 30, 2023 | 50 questions | The exact prompts were not explicitly provided. English was used. |
| Kung et al [62] | ChatGPT-3.5 and ChatGPT-4 with unclear settings | July 2023 | 215 questions | Not clear |
| Lai et al [63] | ChatGPT-4 (May 24; Version 3.5) with unclear settings | Not provided | 200 questions | 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. |
| Lyu et al [64] | ChatGPT with unclear settings | Mid-February 2023 | Not clear | The exact prompts were provided in English. |
| Moise et al [65] | ChatGPT-3.5 with unclear settings | Not provided | 23 questions | The exact prompts were provided in English, with a new session for each question. |
| Oca et al [66] | ChatGPT, Bing, and Bard with unclear settings | April 11, 2023 | 20 queries for each model | The exact prompts were provided in English. |
| Oztermeli & Oztermeli [67] | ChatGPT-3.5 with unclear settings | Not provided | 1177 questions | The exact prompts were not explicitly provided. Turkish was used, with a new session for each question. |
| Pugliese et al [68] | ChatGPT with unclear settings | March 25, 2023 | 15 questions | The exact prompts were provided in English, with a new session for each question. |
| Sallam et al [69] | ChatGPT (default model) with unclear settings | February 25, 2023 | Not provided | The exact prompts were provided in English. |
| Seth et al [70] | ChatGPT-3.5, Bard, and Bing AI | Not provided | 6 questions | The exact prompts were provided in English. |
| Suthar et al [71] | ChatGPT-4 with unclear settings | Not provided | 140 cases | The exact prompts were not explicitly provided. English was used. |
| Walker et al [72] | ChatGPT-4 with unclear settings | Not provided | 5 cases | The exact prompts were not explicitly provided. English was used, with a new session for each question. |
| Wang et al [73] | ChatGPT-3.5 and ChatGPT-4 with unclear settings | February 14, 2023, for ChatGPT-3.5 and May 14-16, 2023, for ChatGPT-4 | 300 questions | The exact prompts were provided. Chinese and English were used. The prompts were enhanced though role play. |
| Wang et al [74] | ChatGPT-3.5 with unclear settings | March 5-10, 2023 | Not clear | Chinese (Mandarin) and English were used. Examples of prompts were provided. |
| Zhou et al [75] | ChatGPT-3.5 with unclear settings | April 24-25, 2023 | Single case and multiple poll questions | The exact prompts were provided in English. |

**Table 2.** Classification based on the evaluation approach of the content generated by the artificial intelligence–based models.

| Authors | Evaluation of performance | Individual role and interrater reliability |
|---|---|---|
| Al-Ashwal et al [42] | Objective via 2 different clinical reference tools | Not applicable |
| Alfertshofer et al [43] | Objective based on the key answers, with the questions screened independently by 4 investigators | Not applicable |
| Ali et al [44] | Objective for multiple-choice questions and true or false questions, and subjective for short-answer and assay questions | Assessment by 2 assessors independently with intraclass correlation coefficient for agreement |
| Aljindan et al [45] | Objective based on key answers and historical performance metrics | Not applicable |
| Altamimi et al [46] | Subjective | Not clear; Assessment for accuracy, informativeness, and accessibility by clinical toxicologists and emergency medicine physicians |
| Baglivo et al [47] | Objective based on key answers and comparison with 5th year medical students' performance | Not applicable |
| Biswas et al [48] | 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 | The median scores across raters for each response were studied; The score represented rater consensus, while the score variance represented disagreements between the raters |
| Chen et al [49] | Objective based on key answers | Not applicable |
| Deiana et al [50] | 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 | Independent assessment by 2 raters with experience in vaccination and health communication topics |
| Fuchs et al [51] | Objective based on key answers | Not applicable |
| Ghosh & Bir [52] | Objective based on key answers; Subjectivity by raters' assessment | 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 |
| Giannos [53] | Objective based on key answers | Not applicable |
| Gobira et al [54] | Objective based on key answers, with an element of subjectivity through classifying the responses as adequate, inadequate, or indeterminate | 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 |
| Grewal et al [55] | Not clear | Not clear |
| Guerra et al [56] | 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 | Not applicable |
| Hamed et al [57] | Subjective | Not clear |
| Hoch et al [58] | Objective based on key answers | Not applicable |
| Juhi et al [59] | Subjective and the use of Stockley's Drug Interactions Pocket Companion 2015 as a reference key | Two raters reached a consensus for categorizing the output |
| Kuang et al [60] | Subjective | Not clear |
| Kumari et al [61] | Subjective; Content validity checked by 2 experts of curriculum design | Three independent raters scored content based on their correctness, with an accuracy score ranging from 1 to 5 |
| Kung et al [62] | Objective based on key answers | Not applicable |
| Lai et al [63] | Objective based on key answers | Not applicable |
| Lyu et al [64] | Subjective | Two experienced radiologists (with 21 and 8 years of experience) evaluated the quality of the ChatGPT responses |
| Moise et al [65] | Subjective through comparison with the latest American Academy of Otolaryngology–Head and Neck Surgery Foundation Clinical Practice Guideline: Tympanostomy Tubes in Children (Update) | 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 |
| Oca et al [66] | Not clear | Not clear |

| Authors | Evaluation of performance | Individual role and interrater reliability |
|---|---|---|
| Oztermeli & Oztermeli [67] | Objective based on key answers | Not applicable |
| Pugliese et al [68] | Subjective using the Likert scale for accuracy, completeness, and comprehensiveness | Multirater: 10 key opinion leaders in nonalcoholic fatty liver disease and 1 nonphysician with expertise in patient advocacy in liver disease independently rating the AI[a] content |
| Sallam et al [69] | Subjective based on correctness, clarity, and conciseness | Fleiss multirater κ |
| Seth et al [70] | 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 | Not clear |
| Suthar et al [71] | Subjective by 3 fellowship-trained neuroradiologists, using a 5-point Likert scale, with 1 indicating "highly improbable" and 5 indicating "highly probable" | Not applicable |
| Walker et al [72] | 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 | 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 |
| Wang et al [73] | Subjective | Unclear |
| Wang et al [74] | Objective based on key answers | Not applicable |
| Zhou et al [75] | Subjective | Unclear |

[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.

| Authors | Transparency | Range | Randomization |
|---|---|---|---|
| Al-Ashwal et al [42] | Full description using 2 tools for assessment: Micromedex, a subscription-based drug-drug interaction screening tool, and Drugs.com, a free database | Narrow; Drug-drug interaction prediction | Nonrandom; Purposeful selection of the drugs by a clinical pharmacist; 5 drugs paired with the top 51 prescribed drugs |
| Alfertshofer et al [43] | 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 | Broad | Randomly extracted |
| Ali et al [44] | 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 | Narrow intersubject (dentistry); Broad intrasubject in restorative dentistry, periodontics, fixed prosthodontics, removable prosthodontics, endodontics, pedodontics, orthodontics, preventive dentistry, oral surgery, and oral medicine | Not clear |
| Aljindan et al [45] | Saudi Medical Licensing Exam questions extracted from the subscription CanadaQBank website | Broad in medicine, with 30% of the questions from medicine, 25% from obstetrics and gynecology, 25% from pediatrics, and the remaining 20% from surgery | Randomized through 4 researchers to ensure comprehensive coverage of questions and eliminate potential bias in question selection |
| Altamimi et al [46] | Snakebite management information guidelines derived from the World Health Organization, Centers for Disease Control and Prevention, and the clinical literature | Narrow | Not clear |
| Baglivo et al [47] | The Italian National Medical Residency test | Narrow; Vaccination-related questions from the Italian National Medical Residency Test | Not clear |
| Biswas et al [48] | Constructed based on the frequently asked questions on the myopia webpage of the Association of British Dispensing Opticians and the College of Optometrists | 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 | Not clear |
| Chen et al [49] | BoardVitals, which is an online question bank accredited by the Accreditation Council for Continuing Medical Education | Neurology-based; Broad intrasubject: basic neuroscience; behavioral, cognitive, psychiatry; cerebrovascular; child neurology; congenital; cranial nerves; critical care; demyelinating 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 | Not clear |
| Deiana et al [50] | Questions concerning vaccine myths and misconceptions by the World Health Organization | Narrow on vaccine myths and misconceptions | Not clear |
| Fuchs et al [51] | Digital platform self-assessment questions tailored for dental and medical students at the University of Bern's Institute for Medical Education | 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 | Not clear |
| Ghosh & Bir [52] | Department question bank, which is a compilation of first and second semester questions from various medical universities across India | Biochemistry | Random without details of randomization |
| Giannos [53] | Specialty Certificate Examination Neurology Web Questions bank | Neurology and neuroscience | Not clear |

| Authors | Transparency | Range | Randomization |
|---|---|---|---|
| Gobira et al [54] | National Brazilian Examination for Revalidation of Medical Diplomas issued by Foreign Higher Education Institutions (Revalida) | Preventive medicine, gynecology and obstetrics, surgery, internal medicine, and pediatrics | Not clear |
| Grewal et al [55] | Not clear | Radiology | Not clear |
| Guerra et al [56] | Questions released by the Congress of Neurological Surgeons in the self-assessment neurosurgery exam | Neurosurgery across 7 subspecialties: tumor, cerebrovascular, trauma, spine, functional, pediatrics, and pain or nerve | Not clear |
| Hamed et al [57] | Guidelines adapted from Diabetes Canada Clinical Practice Guidelines Expert Committee, the Royal Australian College of General Practitioners, Australian Diabetes Society position statement, and the Joint British Diabetes Societies | Management of diabetic ketoacidosis | Not clear |
| Hoch et al [58] | Question database of an online learning platform funded by the German Society of Oto-Rhino-Laryngology, Head and Neck Surgery | 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 | Not clear |
| Juhi et al [59] | A list of drug-drug interactions from previously published research | Narrow on drug-drug interaction | Not clear |
| Kuang et al [60] | Not clear | Neurosurgery | Not clear |
| Kumari et al [61] | Designed by experts in hematology-related cases | Hematology with the following intrasubject aspects: case solving, laboratory calculations, disease interpretations, and other relevant aspects of hematology | Not clear |
| Kung et al [62] | Orthopedic In-Training Examination | Orthopedics | Not clear |
| Lai et al [63] | The United Kingdom Medical Licensing Assessment, which is a newly derived national undergraduate medical exit examination | 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 | Not clear |
| Lyu et al [64] | Chest computed tomography and brain magnetic resonance imaging screening reports collected from the Atrium Health Wake Forest Baptist clinical database | Radiology | Not clear |
| Moise et al [65] | Statements published in the latest American Academy of Otolaryngology–Head and Neck Surgery Foundation Clinical Practice Guideline: Tympanostomy Tubes in Children (Update) | Narrow involving otolaryngology | Not clear |
| Oca et al [66] | Not clear | Narrow involving only queries on accurate recommendation of close ophthalmologists | Not clear |

| Authors | Transparency | Range | Randomization |
|---|---|---|---|
| Oztermeli & Oztermeli [67] | Turkish medical specialty exam, prepared by the Student Selection and Placement Center | 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 | Not clear |
| Pugliese et al [68] | Expert selection of 15 questions commonly asked by patients with nonalcoholic fatty liver disease | Narrow involving nonalcoholic fatty liver disease aspects | Not clear |
| Sallam et al [69] | Panel discussion of experts in health care education | Broad on health care education, medical, dental, pharmacy, and public health | Not clear |
| Seth et al [70] | Devised by 3 fellows of the Royal Australasian College of Surgeons, with experience in performing rhinoplasty and expertise in facial reconstructive surgery | Narrow involving technical aspects of rhinoplasty | Not clear |
| Suthar et al [71] | Quizzes from the Case of the Month feature of the American Journal of Neuroradiology | Narrow involving radiology | Not clear |
| Walker et al [72] | Devised based on the Global Burden of Disease tool | Narrow involving benign and malignant hepatopancreaticobiliary-related conditions | Not clear |
| Wang et al [73] | Medical Exam Help. A total of 10 inpatient and 10 outpatient medical records to form a collection of Chinese medical records after desensitization | Clinical medicine, basic medicine, medical humanities, and relevant laws | Not clear |
| Wang et al [74] | The Taiwanese Senior Professional and Technical Examinations for Pharmacists downloaded from the Ministry of Examination website | 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 | Not clear |
| Zhou et al [75] | A single clinical case from OrthoBullets, a global clinical collaboration platform for orthopedic surgeons; Written permission to use their clinical case report | Very narrow involving a single orthopedic case | Not clear |

## 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.

| Item | Issues considered in each item | Excellent or very good reporting examples |
|---|---|---|
| #1 Model | What is the model of the generative AI[a] tool used for generating content, and what are the exact settings for each tool? | Baglivo et al [47]: Bing, ChatGPT, Chatsonic, Bard, and YouChat, with full details of the mode and large language model, including plugins |
| #2 Evaluation | 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? | 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 assay 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 |
| #3a Timing | When is the generative AI model tested exactly and what are the duration and timing of testing? | Baglivo et al [47]; Biswas et al [48]; Fuchs et al [51]; Ghosh & 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] |
| #3b Transparency | How transparent are the data sources used to generate queries for the generative AI-based model? | Alfertshofer et al [43] |
| #4a Range | What is the range of topics tested and are they intersubject or intrasubject with variability in different subjects? | Ali et al [44]; Chen et al [49]; Hoch et al [58]; and Wang et al [73] |
| #4b Randomization | Was the process of selecting the topics to be tested on the generative AI-based model randomized? | Alfertshofer et al [43] and Aljindan et al [45] |
| #5 Individual | Is there any individual subjective involvement in generative AI content evaluation? If so, did the authors describe the details in full? | Ali et al [44] and Moise et al [65] |
| #6 Count | What is the count of queries executed (sample size)? | Alfertshofer et al [43]; Chen et al [49]; Guerra et al [56]; Hoch et al [58]; and Oztermeli & Oztermeli [67] |
| #7 Specificity of the prompt or language | 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? | 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] |

[a]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 κ interrater reliability ranged from 0.558 to 0.962 ($P$<.001 for the 9 tested items), indicating good to excellent agreement (Table 5).

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) (Table 5).

**Table 5.** Interrater reliability per METRICS item.

| METRICS[a] item | Score | | Quality | Cohen κ | Asymptotic standard error | Approximate T | *P* value |
|---|---|---|---|---|---|---|---|
| | Mean[b] (SD) | Range | | | | | |
| Model | 3.72 (0.58) | 2.5-5.0 | Very good | 0.820 | 0.090 | 6.044 | <.001 |
| Timing | 2.90 (1.93) | 1.0-5.0 | Good | 0.853 | 0.076 | 6.565 | <.001 |
| Count | 3.04 (1.32) | 1.0-5.0 | Good | 0.962 | 0.037 | 10.675 | <.001 |
| Specificity of prompts and language | 3.44 (1.25) | 1.0-5.0 | Very good | 0.765 | 0.086 | 8.083 | <.001 |
| Evaluation | 3.31 (1.16) | 1.0-5.0 | Good | 0.885 | 0.063 | 9.668 | <.001 |
| Individual factors | 2.50 (1.42) | 1.0-5.0 | Satisfactory | 0.865 | 0.087 | 6.860 | <.001 |
| Transparency | 3.24 (1.01) | 1.0-5.0 | Good | 0.558 | 0.112 | 5.375 | <.001 |
| Range | 3.24 (1.07) | 2.0-5.0 | Good | 0.836 | 0.076 | 8.102 | <.001 |
| Randomization | 1.31 (0.87) | 1.0-4.0 | Suboptimal | 0.728 | 0.135 | 5.987 | <.001 |
| Overall | 3.01 (0.58) | 1.5-4.1 | Good | 0.381 | 0.086 | 10.093 | <.001 |

[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


1. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
2. Garg RK, Urs VL, Agarwal AA, Chaudhary SK, Paliwal V, Kar SK. Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review. Health Promot Perspect 2023;13(3):183-191 [FREE Full text] [doi: 10.34172/hpp.2023.22] [Medline: 37808939]
3. Alam F, Lim MA, Zulkipli IN. Integrating AI in medical education: embracing ethical usage and critical understanding. Front Med (Lausanne) 2023;10:1279707 [FREE Full text] [doi: 10.3389/fmed.2023.1279707] [Medline: 37901398]
4. Li J, Dada A, Puladi B, Kleesiek J, Egger J. ChatGPT in healthcare: A taxonomy and systematic review. Comput Methods Programs Biomed 2024 Jan 15;245:108013 [FREE Full text] [doi: 10.1016/j.cmpb.2024.108013] [Medline: 38262126]
5. Mesko B. The ChatGPT (Generative Artificial Intelligence) Revolution Has Made Artificial Intelligence Approachable for Medical Professionals. J Med Internet Res 2023 Jun 22;25:e48392 [FREE Full text] [doi: 10.2196/48392] [Medline: 37347508]
6. Rudolph J, Tan S, Tan S. War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. JALT 2023 Apr 25;6(1):771. [doi: 10.37074/jalt.2023.6.1.23]
7. Choudhury A, Shamszare H. Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. J Med Internet Res 2023 Jun 14;25:e47184 [FREE Full text] [doi: 10.2196/47184] [Medline: 37314848]
8. Shahsavar Y, Choudhury A. User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study. JMIR Hum Factors 2023 May 17;10:e47564 [FREE Full text] [doi: 10.2196/47564] [Medline: 37195756]
9. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J 2021 Jul;8(2):e188-e194 [FREE Full text] [doi: 10.7861/fhj.2021-0095] [Medline: 34286183]
10. Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, et al. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. J Med Internet Res 2023 Aug 22;25:e48659 [FREE Full text] [doi: 10.2196/48659] [Medline: 37606976]
11. Giannos P, Delardas O. Performance of ChatGPT on UK Standardized Admission Tests: Insights From the BMAT, TMUA, LNAT, and TSA Examinations. JMIR Med Educ 2023 Apr 26;9:e47737 [FREE Full text] [doi: 10.2196/47737] [Medline: 37099373]
12. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023 Sep 22;23(1):689 [FREE Full text] [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
13. Miao H, Li C, Wang J. A Future of Smarter Digital Health Empowered by Generative Pretrained Transformer. J Med Internet Res 2023 Sep 26;25:e49963 [FREE Full text] [doi: 10.2196/49963] [Medline: 37751243]
14. Liu T, Xiao X. A Framework of AI-Based Approaches to Improving eHealth Literacy and Combating Infodemic. Front Public Health 2021;9:755808 [FREE Full text] [doi: 10.3389/fpubh.2021.755808] [Medline: 34917575]
15. Choudhury A, Elkefi S, Tounsi A. Exploring factors influencing user perspective of ChatGPT as a technology that assists in healthcare decision making: A cross sectional survey study. medRxiv. 2023. URL: https://www.medrxiv.org/content/10.1101/2023.12.07.23299685v1.full [accessed 2024-02-02]
16. Hsu H, Hsu K, Hou S, Wu C, Hsieh Y, Cheng Y. Examining Real-World Medication Consultations and Drug-Herb Interactions: ChatGPT Performance Evaluation. JMIR Med Educ 2023 Aug 21;9:e48433 [FREE Full text] [doi: 10.2196/48433] [Medline: 37561097]
17. Chen Z. Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanit Soc Sci Commun 2023 Sep 13;10(1):567. [doi: 10.1057/s41599-023-02079-x]
18. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical Considerations of Using ChatGPT in Health Care. J Med Internet Res 2023 Aug 11;25:e48009 [FREE Full text] [doi: 10.2196/48009] [Medline: 37566454]
19. Emsley R. ChatGPT: these are not hallucinations - they're fabrications and falsifications. Schizophrenia (Heidelb) 2023 Aug 19;9(1):52 [FREE Full text] [doi: 10.1038/s41537-023-00379-4] [Medline: 37598184]

20. Gödde D, Nöhl S, Wolf C, Rupert Y, Rimkus L, Ehlers J, et al. A SWOT (Strengths, Weaknesses, Opportunities, and Threats) Analysis of ChatGPT in the Medical Literature: Concise Review. J Med Internet Res 2023 Nov 16;25:e49368 [FREE Full text] [doi: 10.2196/49368] [Medline: 37865883]
21. Kleesiek J, Wu Y, Stiglic G, Egger J, Bian J. An Opinion on ChatGPT in Health Care-Written by Humans Only. J Nucl Med 2023 May 13;64(5):701-703. [doi: 10.2967/jnumed.123.265687] [Medline: 37055219]
22. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics 2021 Sep 15;22(1):122 [FREE Full text] [doi: 10.1186/s12910-021-00687-3] [Medline: 34525993]
23. Mijwil M, Aljanabi M, Ali AH. ChatGPT: Exploring the Role of Cybersecurity in the Protection of Medical Information. MJCS 2023 Feb 1:18-21. [doi: 10.58496/mjcs/2023/004]
24. Sun L, Yin C, Xu Q, Zhao W. Artificial intelligence for healthcare and medical education: a systematic review. Am J Transl Res 2023;15(7):4820-4828 [FREE Full text] [Medline: 37560249]
25. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Med Educ 2023 Feb 08;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
26. Sallam M, Salim NA, Al-Tammemi AB, Barakat M, Fayyad D, Hallit S, et al. ChatGPT Output Regarding Compulsory Vaccination and COVID-19 Vaccine Conspiracy: A Descriptive Study at the Outset of a Paradigm Shift in Online Search for Information. Cureus 2023 Feb;15(2):e35029 [FREE Full text] [doi: 10.7759/cureus.35029] [Medline: 36819954]
27. Scherr R, Halaseh FF, Spina A, Andalib S, Rivera R. ChatGPT Interactive Medical Simulations for Early Clinical Education: Case Study. JMIR Med Educ 2023 Nov 10;9:e49877 [FREE Full text] [doi: 10.2196/49877] [Medline: 37948112]
28. Doshi R, Amin K, Khosla P, Bajaj S, Chheang S, Forman H. Utilizing Large Language Models to Simplify Radiology Reports: a comparative analysis of ChatGPT3.5, ChatGPT4.0, Google Bard, and Microsoft Bing. medRxiv. 2023. URL: https://www.medrxiv.org/content/10.1101/2023.06.04.23290786v2 [accessed 2024-02-02]
29. Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems 2023;3:121-154. [doi: 10.1016/j.iotcps.2023.04.003]
30. Qarajeh A, Tangpanithandee S, Thongprayoon C, Suppadungsuk S, Krisanapan P, Aiumtrakul N, et al. AI-Powered Renal Diet Support: Performance of ChatGPT, Bard AI, and Bing Chat. Clin Pract 2023 Sep 26;13(5):1160-1172 [FREE Full text] [doi: 10.3390/clinpract13050104] [Medline: 37887080]
31. Zúñiga Salazar G, Zúñiga D, Vindel CL, Yoong AM, Hincapie S, Zúñiga A, et al. Efficacy of AI Chats to Determine an Emergency: A Comparison Between OpenAI's ChatGPT, Google Bard, and Microsoft Bing AI Chat. Cureus 2023 Sep;15(9):e45473 [FREE Full text] [doi: 10.7759/cureus.45473] [Medline: 37727841]
32. Fijačko N, Prosen G, Abella BS, Metličar, Štiglic G. Can novel multimodal chatbots such as Bing Chat Enterprise, ChatGPT-4 Pro, and Google Bard correctly interpret electrocardiogram images? Resuscitation 2023 Dec 24;193:110009. [doi: 10.1016/j.resuscitation.2023.110009] [Medline: 37884222]
33. Aiumtrakul N, Thongprayoon C, Suppadungsuk S, Krisanapan P, Miao J, Qureshi F, et al. Navigating the Landscape of Personalized Medicine: The Relevance of ChatGPT, BingChat, and Bard AI in Nephrology Literature Searches. J Pers Med 2023 Sep 30;13(10):1457 [FREE Full text] [doi: 10.3390/jpm13101457] [Medline: 37888068]
34. Stephens LD, Jacobs JW, Adkins BD, Booth GS. Battle of the (Chat)Bots: Comparing Large Language Models to Practice Guidelines for Transfusion-Associated Graft-Versus-Host Disease Prevention. Transfus Med Rev 2023 Jul;37(3):150753 [FREE Full text] [doi: 10.1016/j.tmrv.2023.150753] [Medline: 37704461]
35. Eysenbach G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. JMIR Med Educ 2023 Mar 06;9:e46885 [FREE Full text] [doi: 10.2196/46885] [Medline: 36863937]
36. Meskó B. Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. J Med Internet Res 2023 Oct 04;25:e50638 [FREE Full text] [doi: 10.2196/50638] [Medline: 37792434]
37. Hristidis V, Ruggiano N, Brown EL, Ganta SRR, Stewart S. ChatGPT vs Google for Queries Related to Dementia and Other Cognitive Decline: Comparison of Results. J Med Internet Res 2023 Jul 25;25:e48966 [FREE Full text] [doi: 10.2196/48966] [Medline: 37490317]
38. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
39. Harzing AW. The Publish or Perish Book: Your guide to effective and responsible citation analysis. Melbourne, Australia: Tarma Software Research; 2010.
40. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche P, Vandenbroucke JP, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
41. CASP Qualitative Studies Checklist. Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2023-11-10]

42. Al-Ashwal FY, Zawiah M, Gharaibeh L, Abu-Farha R, Bitar AN. Evaluating the Sensitivity, Specificity, and Accuracy of ChatGPT-3.5, ChatGPT-4, Bing AI, and Bard Against Conventional Drug-Drug Interactions Clinical Tools. Drug Healthc Patient Saf 2023;15:137-147 [FREE Full text] [doi: 10.2147/DHPS.S425858] [Medline: 37750052]
43. Alfertshofer M, Hoch CC, Funk PF, Hollmann K, Wollenberg B, Knoedler S, et al. Sailing the Seven Seas: A Multinational Comparison of ChatGPT's Performance on Medical Licensing Examinations. Ann Biomed Eng 2023 Aug 08:03338-3. [doi: 10.1007/s10439-023-03338-3] [Medline: 37553555]
44. Ali K, Barhom N, Tamimi F, Duggal M. ChatGPT-A double-edged sword for healthcare education? Implications for assessments of dental students. Eur J Dent Educ 2024 Feb 07;28(1):206-211. [doi: 10.1111/eje.12937] [Medline: 37550893]
45. Aljindan FK, Al Qurashi AA, Albalawi IAS, Alanazi AMM, Aljuhani HAM, Falah Almutairi F, et al. ChatGPT Conquers the Saudi Medical Licensing Exam: Exploring the Accuracy of Artificial Intelligence in Medical Knowledge Assessment and Implications for Modern Medical Education. Cureus 2023 Sep;15(9):e45043 [FREE Full text] [doi: 10.7759/cureus.45043] [Medline: 37829968]
46. Altamimi I, Altamimi A, Alhumimidi AS, Altamimi A, Temsah M. Snakebite Advice and Counseling From Artificial Intelligence: An Acute Venomous Snakebite Consultation With ChatGPT. Cureus 2023 Jun;15(6):e40351 [FREE Full text] [doi: 10.7759/cureus.40351] [Medline: 37456381]
47. Baglivo F, De Angelis L, Casigliani V, Arzilli G, Privitera GP, Rizzo C. Exploring the Possible Use of AI Chatbots in Public Health Education: Feasibility Study. JMIR Med Educ 2023 Nov 01;9:e51421 [FREE Full text] [doi: 10.2196/51421] [Medline: 37910155]
48. Biswas S, Logan NS, Davies LN, Sheppard AL, Wolffsohn JS. Assessing the utility of ChatGPT as an artificial intelligence-based large language model for information to answer questions on myopia. Ophthalmic Physiol Opt 2023 Nov 21;43(6):1562-1570. [doi: 10.1111/opo.13207] [Medline: 37476960]
49. Chen TC, Multala E, Kearns P, Delashaw J, Dumont A, Maraganore D, et al. Assessment of ChatGPT's performance on neurology written board examination questions. BMJ Neurol Open 2023;5(2):e000530 [FREE Full text] [doi: 10.1136/bmjno-2023-000530] [Medline: 37936648]
50. Deiana G, Dettori M, Arghittu A, Azara A, Gabutti G, Castiglia P. Artificial Intelligence and Public Health: Evaluating ChatGPT Responses to Vaccination Myths and Misconceptions. Vaccines (Basel) 2023 Jul 07;11(7):1217 [FREE Full text] [doi: 10.3390/vaccines11071217] [Medline: 37515033]
51. Fuchs A, Trachsel T, Weiger R, Eggmann F. ChatGPT's performance in dentistry and allergy-immunology assessments: a comparative study. Swiss Dent J 2023 Oct 06;134(5):1-25 [FREE Full text] [Medline: 37799027]
52. Ghosh A, Bir A. Evaluating ChatGPT's Ability to Solve Higher-Order Questions on the Competency-Based Medical Education Curriculum in Medical Biochemistry. Cureus 2023 Apr;15(4):e37023 [FREE Full text] [doi: 10.7759/cureus.37023] [Medline: 37143631]
53. Giannos P. Evaluating the limits of AI in medical specialisation: ChatGPT's performance on the UK Neurology Specialty Certificate Examination. BMJ Neurol Open 2023;5(1):e000451 [FREE Full text] [doi: 10.1136/bmjno-2023-000451] [Medline: 37337531]
54. Gobira M, Nakayama LF, Moreira R, Andrade E, Regatieri CVS, Belfort R. Performance of ChatGPT-4 in answering questions from the Brazilian National Examination for Medical Degree Revalidation. Rev Assoc Med Bras (1992) 2023;69(10):e20230848 [FREE Full text] [doi: 10.1590/1806-9282.20230848] [Medline: 37792871]
55. Grewal H, Dhillon G, Monga V, Sharma P, Buddhavarapu VS, Sidhu G, et al. Radiology Gets Chatty: The ChatGPT Saga Unfolds. Cureus 2023 Jun;15(6):e40135 [FREE Full text] [doi: 10.7759/cureus.40135] [Medline: 37425598]
56. Guerra GA, Hofmann H, Sobhani S, Hofmann G, Gomez D, Soroudi D, et al. GPT-4 Artificial Intelligence Model Outperforms ChatGPT, Medical Students, and Neurosurgery Residents on Neurosurgery Written Board-Like Questions. World Neurosurg 2023 Nov;179:e160-e165. [doi: 10.1016/j.wneu.2023.08.042] [Medline: 37597659]
57. Hamed E, Eid A, Alberry M. Exploring ChatGPT's Potential in Facilitating Adaptation of Clinical Guidelines: A Case Study of Diabetic Ketoacidosis Guidelines. Cureus 2023 May;15(5):e38784 [FREE Full text] [doi: 10.7759/cureus.38784] [Medline: 37303347]
58. Hoch CC, Wollenberg B, Lüers J, Knoedler S, Knoedler L, Frank K, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. Eur Arch Otorhinolaryngol 2023 Sep;280(9):4271-4278 [FREE Full text] [doi: 10.1007/s00405-023-08051-4] [Medline: 37285018]
59. Juhi A, Pipil N, Santra S, Mondal S, Behera JK, Mondal H. The Capability of ChatGPT in Predicting and Explaining Common Drug-Drug Interactions. Cureus 2023 Mar;15(3):e36272 [FREE Full text] [doi: 10.7759/cureus.36272] [Medline: 37073184]
60. Kuang Y, Zou M, Niu H, Zheng B, Zhang T, Zheng B. ChatGPT encounters multiple opportunities and challenges in neurosurgery. Int J Surg 2023 Oct 01;109(10):2886-2891 [FREE Full text] [doi: 10.1097/JS9.0000000000000571] [Medline: 37352529]
61. Kumari A, Kumari A, Singh A, Singh SK, Juhi A, Dhanvijay AKD, et al. Large Language Models in Hematology Case Solving: A Comparative Study of ChatGPT-3.5, Google Bard, and Microsoft Bing. Cureus 2023 Aug;15(8):e43861 [FREE Full text] [doi: 10.7759/cureus.43861] [Medline: 37736448]

62. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB. Evaluating ChatGPT Performance on the Orthopaedic In-Training Examination. JB JS Open Access 2023;8(3):e23.00056 [FREE Full text] [doi: 10.2106/JBJS.OA.23.00056] [Medline: 37693092]
63. Lai UH, Wu KS, Hsu T, Kan JKC. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. Front Med (Lausanne) 2023 Sep 19;10:1240915 [FREE Full text] [doi: 10.3389/fmed.2023.1240915] [Medline: 37795422]
64. Lyu Q, Tan J, Zapadka ME, Ponnatapura J, Niu C, Myers KJ, et al. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. Vis Comput Ind Biomed Art 2023 May 18;6(1):9 [FREE Full text] [doi: 10.1186/s42492-023-00136-5] [Medline: 37198498]
65. Moise A, Centomo-Bozzo A, Orishchak O, Alnoury MK, Daniel SJ. Can ChatGPT Guide Parents on Tympanostomy Tube Insertion? Children (Basel) 2023 Sep 30;10(10):1634 [FREE Full text] [doi: 10.3390/children10101634] [Medline: 37892297]
66. Oca MC, Meller L, Wilson K, Parikh AO, McCoy A, Chang J, et al. Bias and Inaccuracy in AI Chatbot Ophthalmologist Recommendations. Cureus 2023 Sep;15(9):e45911 [FREE Full text] [doi: 10.7759/cureus.45911] [Medline: 37885556]
67. Oztermeli AD, Oztermeli A. ChatGPT performance in the medical specialty exam: An observational study. Medicine (Baltimore) 2023 Aug 11;102(32):e34673 [FREE Full text] [doi: 10.1097/MD.0000000000034673] [Medline: 37565917]
68. Pugliese N, Wai-Sun Wong V, Schattenberg JM, Romero-Gomez M, Sebastiani G, NAFLD Expert Chatbot Working Group, et al. Accuracy, Reliability, and Comprehensibility of ChatGPT-Generated Medical Responses for Patients With Nonalcoholic Fatty Liver Disease. Clin Gastroenterol Hepatol 2023 Sep 15 (forthcoming) [FREE Full text] [doi: 10.1016/j.cgh.2023.08.033] [Medline: 37716618]
69. Sallam M, Salim N, Barakat M, Al-Tammemi A. ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. Narra J 2023 Mar 29;3(1):e103. [doi: 10.52225/narra.v3i1.103]
70. Seth I, Lim B, Xie Y, Cevik J, Rozen WM, Ross RJ, et al. Comparing the Efficacy of Large Language Models ChatGPT, BARD, and Bing AI in Providing Information on Rhinoplasty: An Observational Study. Aesthet Surg J Open Forum 2023;5:ojad084 [FREE Full text] [doi: 10.1093/asjof/ojad084] [Medline: 37795257]
71. Suthar PP, Kounsal A, Chhetri L, Saini D, Dua SG. Artificial Intelligence (AI) in Radiology: A Deep Dive Into ChatGPT 4.0's Accuracy with the American Journal of Neuroradiology's (AJNR) "Case of the Month". Cureus 2023 Aug;15(8):e43958 [FREE Full text] [doi: 10.7759/cureus.43958] [Medline: 37746411]
72. Walker HL, Ghani S, Kuemmerli C, Nebiker CA, Müller B, Raptis DA, et al. Reliability of Medical Information Provided by ChatGPT: Assessment Against Clinical Guidelines and Patient Information Quality Instrument. J Med Internet Res 2023 Jun 30;25:e47479 [FREE Full text] [doi: 10.2196/47479] [Medline: 37389908]
73. Wang H, Wu W, Dou Z, He L, Yang L. Performance and exploration of ChatGPT in medical examination, records and education in Chinese: Pave the way for medical AI. Int J Med Inform 2023 Sep;177:105173. [doi: 10.1016/j.ijmedinf.2023.105173] [Medline: 37549499]
74. Wang Y, Shen H, Chen T. Performance of ChatGPT on the pharmacist licensing examination in Taiwan. J Chin Med Assoc 2023 Jul 01;86(7):653-658. [doi: 10.1097/JCMA.0000000000000942] [Medline: 37227901]
75. Zhou Y, Moon C, Szatkowski J, Moore D, Stevens J. Evaluating ChatGPT responses in the context of a 53-year-old male with a femoral neck fracture: a qualitative analysis. Eur J Orthop Surg Traumatol 2023 Sep 30 (forthcoming). [doi: 10.1007/s00590-023-03742-4] [Medline: 37776392]
76. Malhotra R, Singh P. Recent advances in deep learning models: a systematic literature review. Multimed Tools Appl 2023 Apr 25;82(29):44977-45060. [doi: 10.1007/s11042-023-15295-z]
77. Hirosawa T, Kawamura R, Harada Y, Mizuta K, Tokumasu K, Kaji Y, et al. ChatGPT-Generated Differential Diagnosis Lists for Complex Case-Derived Clinical Vignettes: Diagnostic Accuracy Evaluation. JMIR Med Inform 2023 Oct 09;11:e48808 [FREE Full text] [doi: 10.2196/48808] [Medline: 37812468]
78. Levkovich I, Elyoseph Z. Suicide Risk Assessments Through the Eyes of ChatGPT-3.5 Versus ChatGPT-4: Vignette Study. JMIR Ment Health 2023 Sep 20;10:e51232 [FREE Full text] [doi: 10.2196/51232] [Medline: 37728984]
79. Flores-Cohaila JA, García-Vicente A, Vizcarra-Jiménez S, De la Cruz-Galán J, Gutiérrez-Arratia J, Quiroga Torres BG, et al. Performance of ChatGPT on the Peruvian National Licensing Medical Examination: Cross-Sectional Study. JMIR Med Educ 2023 Sep 28;9:e48039 [FREE Full text] [doi: 10.2196/48039] [Medline: 37768724]
80. Fraser H, Crossland D, Bacher I, Ranney M, Madsen T, Hilliard R. Comparison of Diagnostic and Triage Accuracy of Ada Health and WebMD Symptom Checkers, ChatGPT, and Physicians for Patients in an Emergency Department: Clinical Data Analysis Study. JMIR Mhealth Uhealth 2023 Oct 03;11:e49995 [FREE Full text] [doi: 10.2196/49995] [Medline: 37788063]
81. Huang RS, Lu KJQ, Meaney C, Kemppainen J, Punnett A, Leung F. Assessment of Resident and AI Chatbot Performance on the University of Toronto Family Medicine Residency Progress Test: Comparative Study. JMIR Med Educ 2023 Sep 19;9:e50514 [FREE Full text] [doi: 10.2196/50514] [Medline: 37725411]
82. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 2021 Mar 31;8(1):53 [FREE Full text] [doi: 10.1186/s40537-021-00444-8] [Medline: 33816053]

83. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Information Fusion 2023 Nov;99:101805. [doi: 10.1016/j.inffus.2023.101805]
84. Giray L. Prompt Engineering with ChatGPT: A Guide for Academic Writers. Ann Biomed Eng 2023 Dec;51(12):2629-2633. [doi: 10.1007/s10439-023-03272-4] [Medline: 37284994]
85. Khlaif ZN, Mousa A, Hattab MK, Itmazi J, Hassan AA, Sanmugam M, et al. The Potential and Concerns of Using AI in Scientific Research: ChatGPT Performance Evaluation. JMIR Med Educ 2023 Sep 14;9:e47049 [FREE Full text] [doi: 10.2196/47049] [Medline: 37707884]
86. Varsha PS. How can we manage biases in artificial intelligence systems – A systematic literature review. International Journal of Information Management Data Insights 2023 Apr;3(1):100165. [doi: 10.1016/j.jjimei.2023.100165]
87. Yanagita Y, Yokokawa D, Uchida S, Tawara J, Ikusaka M. Accuracy of ChatGPT on Medical Questions in the National Medical Licensing Examination in Japan: Evaluation Study. JMIR Form Res 2023 Oct 13;7:e48023 [FREE Full text] [doi: 10.2196/48023] [Medline: 37831496]
88. Taira K, Itaya T, Hanada A. Performance of the Large Language Model ChatGPT on the National Nurse Examinations in Japan: Evaluation Study. JMIR Nurs 2023 Jun 27;6:e47305 [FREE Full text] [doi: 10.2196/47305] [Medline: 37368470]
89. Sallam M, Barakat M, Sallam M. Pilot Testing of a Tool to Standardize the Assessment of the Quality of Health Information Generated by Artificial Intelligence-Based Models. Cureus 2023 Nov;15(11):e49373 [FREE Full text] [doi: 10.7759/cureus.49373] [Medline: 38024074]
90. Sezgin E, Chekeni F, Lee J, Keim S. Clinical Accuracy of Large Language Models and Google Search Responses to Postpartum Depression Questions: Cross-Sectional Study. J Med Internet Res 2023 Sep 11;25:e49240 [FREE Full text] [doi: 10.2196/49240] [Medline: 37695668]
91. Wilhelm TI, Roos J, Kaczmarczyk R. Large Language Models for Therapy Recommendations Across 3 Clinical Specialties: Comparative Study. J Med Internet Res 2023 Oct 30;25:e49324 [FREE Full text] [doi: 10.2196/49324] [Medline: 37902826]
92. Ferreira AL, Chu B, Grant-Kels JM, Ogunleye T, Lipoff JB. Evaluation of ChatGPT Dermatology Responses to Common Patient Queries. JMIR Dermatol 2023 Nov 17;6:e49280 [FREE Full text] [doi: 10.2196/49280] [Medline: 37976093]
93. Thirunavukarasu AJ, Hassan R, Mahmood S, Sanghera R, Barzangi K, El Mukashfi M, et al. Trialling a Large Language Model (ChatGPT) in General Practice With the Applied Knowledge Test: Observational Study Demonstrating Opportunities and Limitations in Primary Care. JMIR Med Educ 2023 Apr 21;9:e46599 [FREE Full text] [doi: 10.2196/46599] [Medline: 37083633]
94. Lakdawala N, Channa L, Gronbeck C, Lakdawala N, Weston G, Sloan B, et al. Assessing the Accuracy and Comprehensiveness of ChatGPT in Offering Clinical Guidance for Atopic Dermatitis and Acne Vulgaris. JMIR Dermatol 2023 Nov 14;6:e50409 [FREE Full text] [doi: 10.2196/50409] [Medline: 37962920]
95. Borchert RJ, Hickman CR, Pepys J, Sadler TJ. Performance of ChatGPT on the Situational Judgement Test-A Professional Dilemmas-Based Examination for Doctors in the United Kingdom. JMIR Med Educ 2023 Aug 07;9:e48978 [FREE Full text] [doi: 10.2196/48978] [Medline: 37548997]
96. Sun H, Zhang K, Lan W, Gu Q, Jiang G, Yang X, et al. An AI Dietitian for Type 2 Diabetes Mellitus Management Based on Large Language and Image Recognition Models: Preclinical Concept Validation Study. J Med Internet Res 2023 Nov 09;25:e51300 [FREE Full text] [doi: 10.2196/51300] [Medline: 37943581]
97. Ettman CK, Galea S. The Potential Influence of AI on Population Mental Health. JMIR Ment Health 2023 Nov 16;10:e49936 [FREE Full text] [doi: 10.2196/49936] [Medline: 37971803]
98. Sallam M, Barakat M, Sallam M. METRICS: Establishing a Preliminary Checklist to Standardize the Design and Reporting of Generative Artificial Intelligence-Based Studies in Healthcare Education and Practice. Interact J Med Res 2024 Jan 26:54704 [FREE Full text] [doi: 10.2196/54704] [Medline: 38276872]
99. Sallam M, Al-Salahat K. Below average ChatGPT performance in medical microbiology exam compared to university students. Front. Educ 2023 Dec 21;8:1333415. [doi: 10.3389/feduc.2023.1333415]
100. Sallam M, Al-Salahat K, Al-Ajlouni E. ChatGPT Performance in Diagnostic Clinical Microbiology Laboratory-Oriented Case Scenarios. Cureus 2023 Dec;15(12):e50629 [FREE Full text] [doi: 10.7759/cureus.50629] [Medline: 38107211]


## Abbreviations

**AI:** artificial intelligence
**CASP:** Critical Appraisal Skills Programme
**METRICS:** Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, and Specificity of prompts and language

*Edited by T de Azevedo Cardoso; submitted 19.11.23; peer-reviewed by I Albalawi, A Spina, F Halaseh; comments to author 13.12.23; revised version received 18.12.23; accepted 26.01.24; published 15.02.24.*

*Please cite as:*
*Sallam M, Barakat M, Sallam M*
*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*
*Interact J Med Res 2024;13:e54704*
*URL: https://www.i-jmr.org/2024/1/e54704*
*doi:10.2196/54704*
*PMID:38276872*

Original Paper

# Research Trends and Evolution in Radiogenomics (2005-2023): Bibliometric Analysis

Meng Wang[1], MM; Yun Peng[2], MM; Ya Wang[2], BMed; Dehong Luo[1], MD

[1]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, Shenzhen, China

[2]Department 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 Baohe 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.

***(Interact J Med Res 2024;13:e51347)*** doi:10.2196/51347

**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, radigenomics 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 language and document type of the retrieved documents sequentially. The exclusion criteria were as follows: (1) document type is not a research or review article (eg, proceeding paper, meeting abstract, or editorial) and (2) articles written in languages other than English (eg, Japanese or Chinese). There were 420 papers excluded due to an inappropriate document type and 12 papers excluded due to language; thus, 3237 papers remained for analysis, which were exported in plain-text format. The information on corresponding countries, institutions, authors, and journals obtained from the WoSCC were recorded. The original data of the retrieved articles are provided in Multimedia Appendix 1.

### Data Analysis and Visualization

After detecting duplicate documents using CiteSpace (version 6.3.R1), Excel 2016, CiteSpace, and VOSviewer (version 1.6.19.0) were used to perform the bibliometrics analysis. Excel was used to analyze the annual publications, whereas VOSviewer was used to generate visualization networks of keywords co-occurrence and of collaboration among countries and institutions. The networks were constructed with keywords, countries, and institutions as nodes, respectively. The thickness of the line connecting nodes indicates the strength of the association. The betweenness centrality (BC) parameter was used to assess the importance of each node in the network; a higher BC value signifies greater importance in the network [17].

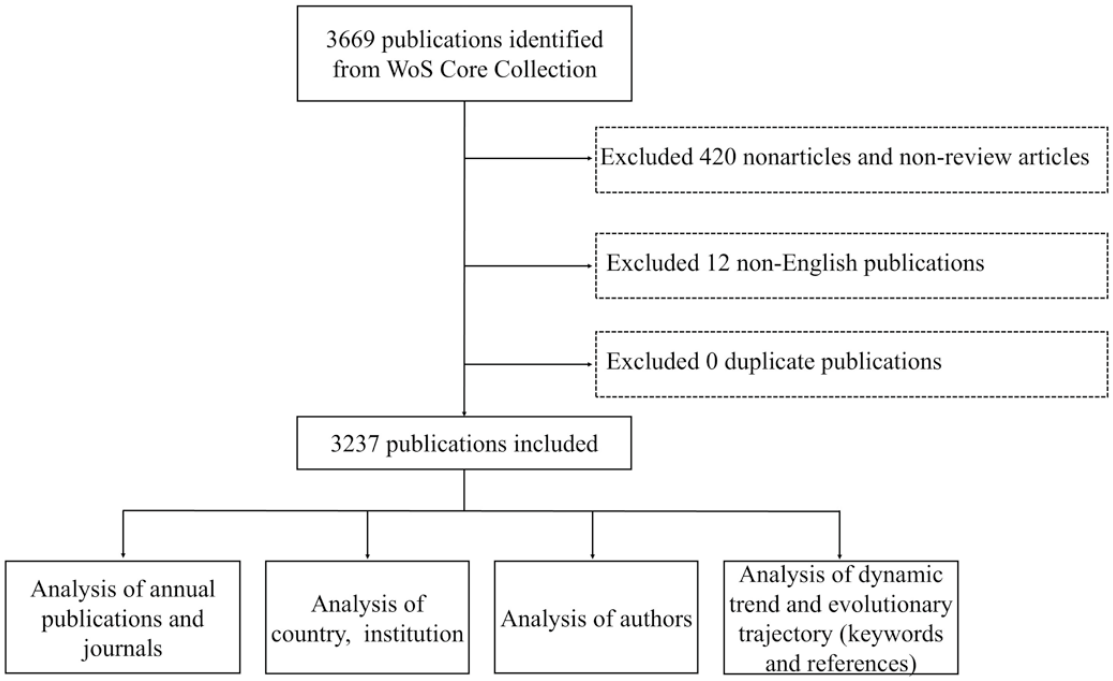
CiteSpace was used to determine the citation burst of keywords and to map the timeline of references by keywords. Burst detection can detect the citation burst of a specific keyword (or document) within a certain field, at least within a short period of time during a given time frame [18]. Keywords with a burst of occurrences are indicators of hot spot research topics in a field. According to the instructions of CiteSpace [19], the timeline of reference clusters is an analysis based on references for the exported literature. CiteSpace can extract noun phrases from the titles, keyword lists, or abstracts of articles that cited the particular cluster. The automatically selected labels will be displayed and the clusters are numbered in descending order of the cluster size, starting from the largest as cluster 0, the second largest as cluster 1, and so on. In this way, the network characterizes the development of the field over time, showing the most important footprints of the related research activities.

Certain data, including the number of publications, impact factor, h-index, and average per item of journals, countries, and authors, were retrieved from the WoSCC website. The h-index and average per item were used in the country, institution, and author analyses. The h-index was introduced by Hirsch [20] in 2005 and is commonly used as a scientific contribution metric corresponding to the number of times a paper is cited. The average per item is calculated by dividing the total number of citations by the number of publications, resulting in the average number of citations per publication.

### 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.

**Figure 1.** Workflow of the analytical procedures. WoS: Web of Science.

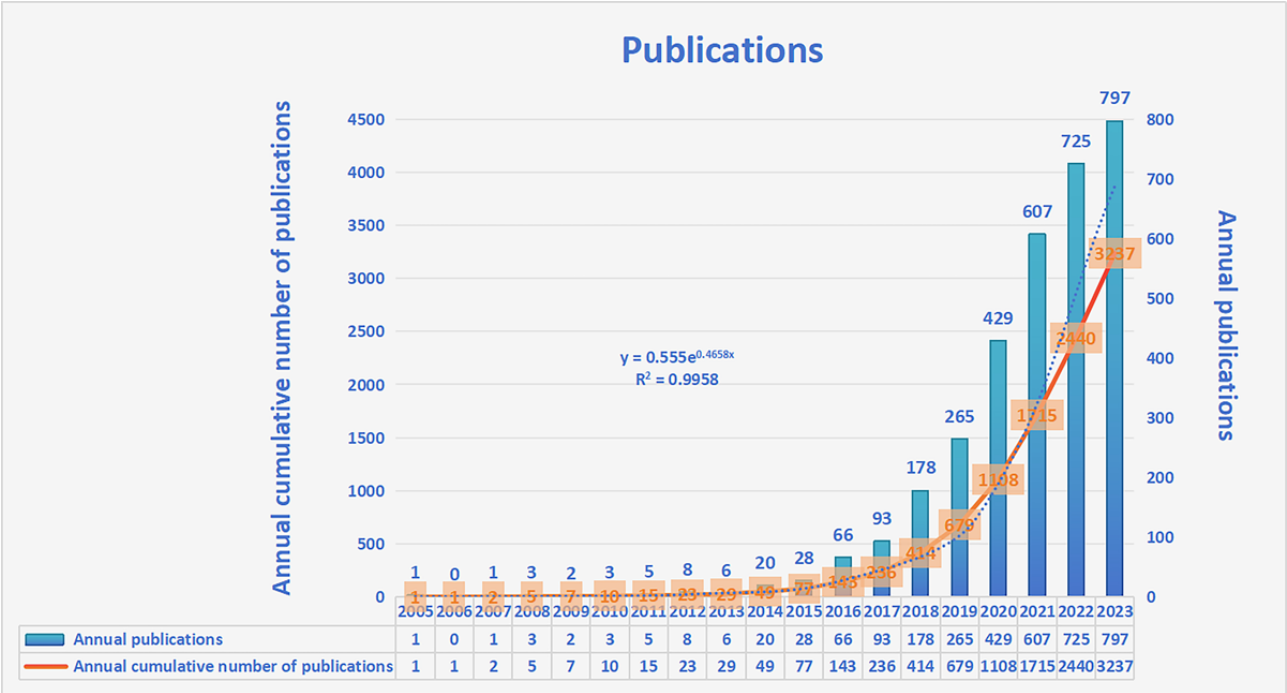



# Results

## 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).

**Figure 2.** Global trend of publications on radiogenomics from 2005 to 2023.




## 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.

| Rank | Journal | Articles, n | IF[a] (2022) |
|---|---|---|---|
| 1 | Frontiers in Oncology | 284 | 4.7003 |
| 2 | Cancers | 196 | 5.5999 |
| 3 | European Radiology | 135 | 5.9003 |
| 4 | Scientific Reports | 103 | 4.6 |
| 5 | Diagnostics | 73 | 3.5999 |
| 6 | Medical Physics | 63 | 3.8001 |
| 7 | Journal of Magnetic Resonance Imaging | 60 | 4.3997 |
| 8 | European Journal of Radiology | 57 | 3.3 |
| 9 | Academic Radiology | 50 | 4.8003 |
| 10 | Physics in Medicine And Biology | 49 | 3.5 |
| 11 | British Journal of Radiology | 45 | 2.6002 |
| 12 | Radiology | 41 | 19.7005 |
| 13 | European Journal of Nuclear Medicine and Molecular Imaging | 40 | 9.1005 |
| 14 | Abdominal Radiology | 34 | 2.4002 |
| 15 | BMC Medical Imaging | 33 | 2.7001 |

[a]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).

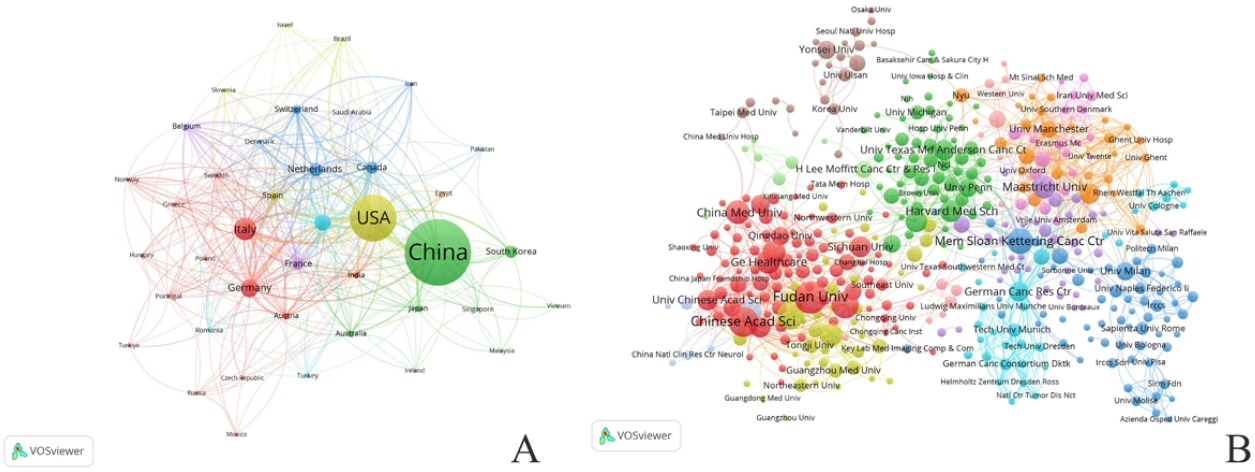
**Table 2.** Top 10 productive countries with the most publications in the field of radiogenomics.

| Rank | Total link strength | BC[a] | Country | Publications, n | H-index | Times cited | Average per item |
|---|---|---|---|---|---|---|---|
| 1 | 329 | 0.11 | China | 1470 | 68 | 25,319 | 17.22 |
| 2 | 820 | 0.36 | United States | 891 | 87 | 46,426 | 52.11 |
| 3 | 311 | 0.13 | Italy | 326 | 41 | 7327 | 22.48 |
| 4 | 353 | 0.06 | Germany | 233 | 39 | 7966 | 34.19 |
| 5 | 418 | 0.19 | England | 209 | 37 | 6185 | 29.59 |
| 6 | 328 | 0.1 | Netherlands | 161 | 37 | 16,077 | 99.86 |
| 7 | 54 | 0.01 | South Korea | 142 | 31 | 4296 | 30.25 |
| 8 | 248 | 0.12 | Canada | 140 | 36 | 7204 | 51.46 |
| 9 | 223 | 0.08 | France | 128 | 32 | 4099 | 32.02 |
| 10 | 181 | 0.01 | Switzerland | 88 | 23 | 1893 | 21.51 |

[a]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.

| Rank | Publications | | | | | Citations | | |
|---|---|---|---|---|---|---|---|---|
| | Author | Articles, n | Country | H-index | Average per item | Author | Citations, n | Country |
| 1 | Philippe Lambin | 33 | Netherlands | 21 | 358.06 | Philippe Lambin | 1573 | Netherlands |
| 2 | Catharine M West | 32 | England | 17 | 44.78 | Robyn Gillies | 1109 | Australia |
| 3 | Robyn Gillies | 32 | Australia | 24 | 486.13 | Hugo Aerts | 965 | United States |
| 4 | Evis Sala | 26 | Italy | 17 | 35.92 | JJM Van Griethuy-sen | 720 | United States |
| 5 | Andre Dekker | 25 | Netherlands | 12 | 448.4 | Alex Zwanenburg | 597 | Germany |
| 6 | Jie Tian | 25 | China | 14 | 48.04 | Chintan Parmar | 448 | United States |
| 7 | Sarah L Kerns | 25 | United States | 17 | 44.72 | David N Louis | 429 | United States |
| 8 | Hugo Aerts | 24 | United States | 21 | 526.42 | Philip Kickingered-er | 390 | Germany |
| 9 | Shaofeng Duan, | 24 | China | 11 | 12.67 | Yan-Qi Huang | 382 | China |
| 10 | Seung-Koo Lee | 22 | South Korea | 10 | 23.91 | Kumar Vinod | 376 | India |

## 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.

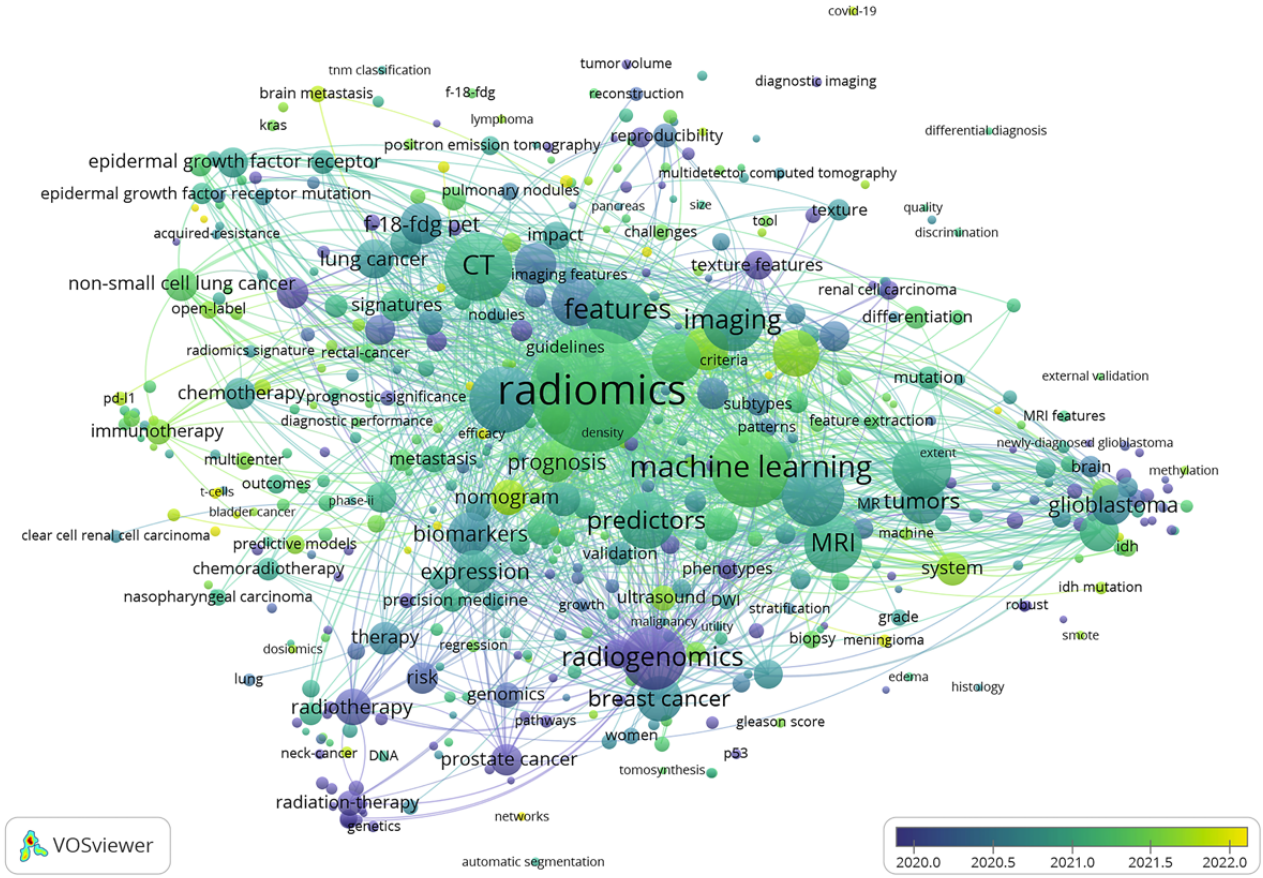


**Table 4.** The top 30 keywords with the highest frequency in the field of radiogenomics.

| Rank | Keywords | Total link strength | Frequency |
|---|---|---|---|
| 1 | radiomics | 13,538 | 1918 |
| 2 | machine learning | 5816 | 779 |
| 3 | CT[a] | 4300 | 580 |
| 4 | carcinoma | 4250 | 569 |
| 5 | features | 4279 | 538 |
| 6 | radiogenomics | 4102 | 535 |
| 7 | survival analysis | 3984 | 500 |
| 8 | imaging | 3650 | 478 |
| 9 | classification | 3367 | 439 |
| 10 | MRI[b] | 3323 | 423 |
| 11 | predictors | 3427 | 416 |
| 12 | diagnosis | 2247 | 310 |
| 13 | texture analysis | 2489 | 303 |
| 14 | deep learning | 2076 | 273 |
| 15 | tumors | 2095 | 271 |
| 16 | breast cancer | 1958 | 260 |
| 17 | artificial intelligence | 2074 | 254 |
| 18 | expression | 1906 | 242 |
| 19 | biomarkers | 1906 | 240 |
| 20 | heterogeneity | 1902 | 227 |
| 21 | F-18-FDG PET[c] | 1874 | 225 |
| 22 | prognosis | 1670 | 219 |
| 23 | glioblastoma | 1806 | 212 |
| 24 | gliomas | 1611 | 198 |
| 25 | lung cancer | 1485 | 193 |
| 26 | radiotherapy | 1258 | 175 |
| 27 | nomogram | 1228 | 167 |
| 28 | models | 1083 | 152 |
| 29 | recurrence | 1329 | 152 |
| 30 | system | 983 | 144 |

[a]CT: computed tomography.

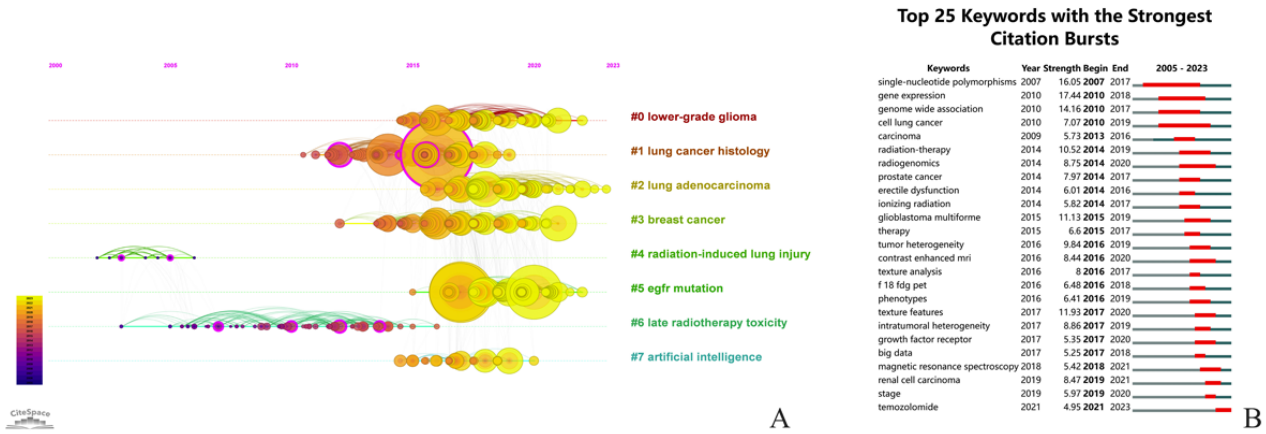
[b]MRI: magnetic resonance imaging.

[c]F-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.

**Figure 5.** Timeline view of reference clustering analysis on radiogenomics (A) and top 25 keywords with the strongest citation bursts (B).




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” (fludeoxyglucose 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.”

## Discussion

### Principal Findings

The concept of precision medicine has propelled increased attention toward radiogenomics, a fusion of genomics and radiomics, to achieve personalized treatment, owing to its potential as a noninvasive tool to predict treatment responses. This study analyzed 3237 relevant documents in the field of radiogenomics published between 2005 and 2023 from the WoSCC. The increasing number of annual publications, especially the extremely high growth rate after 2017, indicates how interest in radiogenomics research in the clinical field has been increasing from year to year.

### Current Status of Publications for Countries and Authors

China currently has the highest number of publications in radiogenomics, although the total citation count is lower than that for the United States, with the average citation number per item lower than that of the other top 10 countries. The Netherlands, the United States, and Canada obtained the highest averages per item. It can be concluded that the United States and the Netherlands have performed reasonably well both in terms of the number and quality of published documents, demonstrating their strong influence in the field. Philippe Lambin of the Netherlands ranked first in both number of publications and number of citations, indicating his major contributions to this field.

### 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 of lesions 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 radigenomics 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

1. Moussa AM, Ziv E. Radiogenomics in interventional oncology. Curr Oncol Rep 2021 Jan 02;23(1):9. [doi: 10.1007/s11912-020-00994-9] [Medline: 33387095]
2. Fan M, Xia P, Clarke R, Wang Y, Li L. Radiogenomic signatures reveal multiscale intratumour heterogeneity associated with biological functions and survival in breast cancer. Nat Commun 2020 Sep 25;11(1):4861. [doi: 10.1038/s41467-020-18703-2] [Medline: 32978398]
3. Mazurowski MA. Radiogenomics: what it is and why it is important. J Am Coll Radiol 2015 Aug;12(8):862-866. [doi: 10.1016/j.jacr.2015.04.019] [Medline: 26250979]
4. Badic B, Tixier F, Cheze Le Rest C, Hatt M, Visvikis D. Radiogenomics in colorectal cancer. Cancers 2021 Feb 26;13(5):973 [FREE Full text] [doi: 10.3390/cancers13050973] [Medline: 33652647]
5. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic features are associated with EGFR mutation status in lung adenocarcinomas. Clin Lung Cancer 2016 Sep;17(5):441-448 [FREE Full text] [doi: 10.1016/j.cllc.2016.02.001] [Medline: 27017476]
6. Zhu X, Lazow MA, Schafer A, Bartlett A, Senthil Kumar S, Mishra DK, et al. A pilot radiogenomic study of DIPG reveals distinct subgroups with unique clinical trajectories and therapeutic targets. Acta Neuropathol Commun 2021 Jan 11;9(1):14 [FREE Full text] [doi: 10.1186/s40478-020-01107-0] [Medline: 33431066]
7. Patel SH, Poisson LM, Brat DJ, Zhou Y, Cooper L, Snuderl M, et al. T2-FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: a TCGA/TCIA Project. Clin Cancer Res 2017 Oct 15;23(20):6078-6085. [doi: 10.1158/1078-0432.CCR-17-0560] [Medline: 28751449]
8. Perez-Johnston R, Araujo-Filho JA, Connolly JG, Caso R, Whiting K, Tan KS, et al. CT-based radiogenomic analysis of clinical stage I lung adenocarcinoma with histopathologic features and oncologic outcomes. Radiology 2022 Jun;303(3):664-672 [FREE Full text] [doi: 10.1148/radiol.211582] [Medline: 35230187]
9. Park EK, Lee K, Seo BK, Cho KR, Woo OH, Son GS, et al. Machine learning approaches to radiogenomics of breast cancer using low-dose perfusion computed tomography: predicting prognostic biomarkers and molecular subtypes. Sci Rep 2019 Nov 28;9(1):17847. [doi: 10.1038/s41598-019-54371-z] [Medline: 31780739]
10. Kickingereder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, et al. Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. Radiology 2016 Dec;281(3):907-918. [doi: 10.1148/radiol.2016161382] [Medline: 27636026]
11. Infante T, Cavaliere C, Punzo B, Grimaldi V, Salvatore M, Napoli C. Radiogenomics and artificial intelligence approaches applied to cardiac computed tomography angiography and cardiac magnetic resonance for precision medicine in coronary heart disease: a systematic review. Circ: Cardiovascular Imaging 2021 Dec;14(12):1133-1146. [doi: 10.1161/circimaging.121.013025]
12. El Naqa I, Kerns SL, Coates J, Luo Y, Speers C, West CML, et al. Radiogenomics and radiotherapy response modeling. Phys Med Biol 2017 Aug 01;62(16):R179-R206 [FREE Full text] [doi: 10.1088/1361-6560/aa7c55] [Medline: 28657906]
13. Liu Z, Duan T, Zhang Y, Weng S, Xu H, Ren Y, et al. Radiogenomics: a key component of precision cancer medicine. Br J Cancer 2023 Sep;129(5):741-753. [doi: 10.1038/s41416-023-02317-8] [Medline: 37414827]
14. Wang Y, Peng Y, Wang T, Li H, Zhao Z, Gong L, et al. The evolution and current situation in the application of dual-energy computed tomography: a bibliometric study. Quant Imaging Med Surg 2023 Oct 01;13(10):6801-6813 [FREE Full text] [doi: 10.21037/qims-23-467] [Medline: 37869341]
15. Peng Y, Wang T, Wang J, Wang H, Fang R, Gong L, et al. The application of artificial intelligence in thyroid nodules: a systematic review based on bibliometric analysis. Endocr Metab Immune Disord Drug Targets 2024 Jan 04:online ahead of print. [doi: 10.2174/0118715303264254231117113456] [Medline: 38178659]
16. West CML, McKay MJ, Hölscher T, Baumann M, Stratford IJ, Bristow RG, et al. Molecular markers predicting radiotherapy response: report and recommendations from an International Atomic Energy Agency technical meeting. Int J Radiat Oncol Biol Phys 2005 Aug 01;62(5):1264-1273. [doi: 10.1016/j.ijrobp.2005.05.001] [Medline: 16029781]
17. Zhao X, Liu Z. Analysis of topological parameters of complex disease genes reveals the importance of location in a biomolecular network. Genes 2019 Feb 14;10(2):143 [FREE Full text] [doi: 10.3390/genes10020143] [Medline: 30769902]
18. Gan P, Huang S, Pan X, Xia H, Zeng X, Ren W, et al. Global research trends in the field of liver cirrhosis from 2011 to 2020: a visualised and bibliometric study. World J Gastroenterol 2022 Sep 07;28(33):4909-4919 [FREE Full text] [doi: 10.3748/wjg.v28.i33.4909] [Medline: 36156929]
19. Chen C. How to Use CiteSpace. Victoria, BC: LeanPub; 2022.
20. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005 Nov 15;102(46):16569-16572 [FREE Full text] [doi: 10.1073/pnas.0507655102] [Medline: 16275915]
21. Rindi G, Klimstra DS, Abedi-Ardekani B, Asa SL, Bosman FT, Brambilla E, et al. A common classification framework for neuroendocrine neoplasms: an International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. Mod Pathol 2018 Dec;31(12):1770-1786 [FREE Full text] [doi: 10.1038/s41379-018-0110-y] [Medline: 30140036]
22. Zhao X, Xiang Z, Chen Y, Yang P, Hu Y, Zeng Z. A sequence polymorphism on 8q24 is associated with survival in hepatocellular carcinoma patients who received radiation therapy. Sci Rep 2018 Feb 02;8(1):2264. [doi: 10.1038/s41598-018-20700-x] [Medline: 29396413]

23. Jiang W, Han X, Wang J, Wang L, Xu Z, Wei Q, et al. miR-22 enhances the radiosensitivity of small-cell lung cancer by targeting the WRNIP1. J Cell Biochem 2019 Oct 12;120(10):17650-17661 [FREE Full text] [doi: 10.1002/jcb.29032] [Medline: 31190355]
24. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007 May 1;39(5):631-637. [doi: 10.1038/ng1999] [Medline: 17401366]
25. Lhuillier C, Rudqvist N, Elemento O, Formenti SC, Demaria S. Radiation therapy and anti-tumor immunity: exposing immunogenic mutations to the immune system. Genome Med 2019 Jun 20;11(1):40 [FREE Full text] [doi: 10.1186/s13073-019-0653-7] [Medline: 31221199]
26. Peeken JC, Bernhofer M, Wiestler B, Goldberg T, Cremers D, Rost B, et al. Radiomics in radiooncology - challenging the medical physicist. Phys Med 2018 Apr;48:27-36. [doi: 10.1016/j.ejmp.2018.03.012] [Medline: 29728226]
27. Doshi AM, Ream JM, Kierans AS, Bilbily M, Rusinek H, Huang WC, et al. Use of MRI in differentiation of papillary renal cell carcinoma subtypes: qualitative and quantitative analysis. AJR Am J Roentgenol 2016 Mar;206(3):566-572. [doi: 10.2214/AJR.15.15004] [Medline: 26901013]
28. Yeh AC, Li H, Zhu Y, Zhang J, Khramtsova G, Drukker K, et al. Radiogenomics of breast cancer using dynamic contrast enhanced MRI and gene expression profiling. Cancer Imaging 2019 Jul 15;19(1):48 [FREE Full text] [doi: 10.1186/s40644-019-0233-5] [Medline: 31307537]
29. LoGullo R, Horvat J, Reiner J, Pinker K. Multimodal, multiparametric and genetic breast imaging. Radiologe 2021 Feb;61(2):183-191. [doi: 10.1007/s00117-020-00801-3] [Medline: 33464404]
30. Pinker K, Chin J, Melsaether AN, Morris EA, Moy L. Precision medicine and radiogenomics in breast cancer: new approaches toward diagnosis and treatment. Radiology 2018 Jun;287(3):732-747. [doi: 10.1148/radiol.2018172171] [Medline: 29782246]


## 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


*Edited by T de Azevedo Cardoso; submitted 28.07.23; peer-reviewed by LV Ivanitskaya, A Banerjee; comments to author 15.11.23; revised version received 10.03.24; accepted 20.05.24; published 09.07.24.*

*Please cite as:*
*Wang M, Peng Y, Wang Y, Luo D*
*Research Trends and Evolution in Radiogenomics (2005-2023): Bibliometric Analysis*
*Interact J Med Res 2024;13:e51347*
*URL: https://www.i-jmr.org/2024/1/e51347*
*doi:10.2196/51347*
*PMID:38980713*

Original Paper

# Knowledge, Attitudes, and Behaviors Toward Salt Consumption and Its Association With 24-Hour Urinary Sodium and Potassium Excretion in Adults Living in Mexico City: Cross-Sectional Study

Gabriela Gutiérrez-Salmeán[1,2*], PhD; Paola Vanessa Miranda-Alatriste[3*], PhD; Patricio Benítez-Alday[3], MD; Luis Enrique Orozco-Rivera[2], MSc; Nurit Islas-Vargas[2], BSc; Ángeles Espinosa-Cuevas[3], PhD; Ricardo Correa-Rotter[3], MD; Eloisa Colin-Ramirez[1,4], PhD

[1]Centro de Investigación en Ciencias de la Salud, Facultad de Ciencias de la Salud, Universidad Anáhuac México, Huixquilucan, Estado de México, Mexico

[2]Servicio de Nutrición, Centro de Especialidades del Riñón, Ciudad de México, Mexico

[3]Departamento de Nefrología y Metabolismo Mineral, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Ciudad de México, Mexico

[4]Dirección de Nutrición, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Ciudad de México, Mexico

[*]these authors contributed equally

**Corresponding Author:**
Eloisa Colin-Ramirez, PhD
Centro de Investigación en Ciencias de la Salud
Facultad de Ciencias de la Salud
Universidad Anáhuac México
Avenida Universidad Anáhuac 46,
Lomas Anáhuac
Huixquilucan, Estado de México, 52786
Mexico
Phone: 52 55 5627 0210
Email: eloisa.colinr@incmnsz.mx

## Abstract

**Background:** The World Health Organization recommends a daily sodium intake of less than 2000 mg for adults; however, the Mexican population, like many others globally, consumes more sodium than this recommended amount. Excessive sodium intake is often accompanied by inadequate potassium intake. The association between knowledge, attitudes, and behaviors (KAB) and actual sodium intake has yielded mixed results across various populations. In Mexico, however, salt/sodium-related KAB and its relationship with sodium and potassium intake have not been evaluated.

**Objective:** This study primarily aims to describe salt/sodium-related KAB in a Mexican population and, secondarily, to explore the association between KAB and 24-hour urinary sodium and potassium excretion.

**Methods:** We conducted a cross-sectional study in an adult population from Mexico City and the surrounding metropolitan area. Self-reported KAB related to salt/sodium intake was assessed using a survey developed by the Pan American Health Organization. Anthropometric measurements were taken, and 24-hour urinary sodium and potassium excretion levels were determined. Descriptive statistics were stratified by sex and presented as means (SD) or median (25th-75th percentiles) for continuous variables, and as absolute and relative frequencies for categorical variables. The associations between KAB and sodium and potassium excretion were assessed using analysis of covariance, adjusting for age, sex, BMI, and daily energy intake as covariates, with the Šidák correction applied for multiple comparisons.

**Results:** Overall, 232 participants were recruited (women, n=184, 79.3%). The mean urinary sodium and potassium excretion were estimated to be 2582.5 and 1493.5 mg/day, respectively. A higher proportion of men did not know the amount of sodium they consumed compared with women (12/48, 25%, vs 15/184, 8.2%, P=.01). More women reported knowing that there is a recommended amount for daily sodium intake than men (46/184, 25%, vs 10/48, 20.8%, P=.02). Additionally, more than half of men (30/48, 62.5%) reported never or rarely reading food labels, compared with women (96/184, 52.1%, P=.04). Better salt/sodium-related KAB was associated with higher adjusted mean sodium and potassium excretion. For example, mean sodium excretion was 3011.5 (95% CI 2640.1-3382.9) mg/day among participants who reported knowing the difference between salt and sodium, compared with 2592.8 (95% CI 2417.2-2768.3) mg/day in those who reported not knowing this difference (P=.049).

Similarly, potassium excretion was 1864.9 (95% CI 1669.6-2060.3) mg/day for those who knew the difference, compared with 1512.5 (95% CI 1420.1-1604.8) mg/day for those who did not (P=.002). Additionally, higher urinary sodium excretion was observed among participants who reported consuming too much sodium (3216.0 mg/day, 95% CI 2867.1-3565.0 mg/day) compared with those who claimed to eat just the right amount (2584.3 mg/day, 95% CI 2384.9-2783.7 mg/day, P=.01).

**Conclusions:** Salt/sodium-related KAB was poor in this study sample. Moreover, KAB had a greater impact on potassium excretion than on sodium excretion, highlighting the need for more strategies to improve KAB related to salt/sodium intake. Additionally, it is important to consider other strategies aimed at modifying the sodium content of foods.

***(Interact J Med Res 2024;13:e57265)*** doi:10.2196/57265

**KEYWORDS**

beliefs; attitudes; hypertension; knowledge; salt consumption; sodium intake; potassium intake; Mexico

## Introduction

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, with hypertension being one of the primary contributors, affecting over 20% of the Mexican adult population [1]. Excessive sodium intake is recognized as a major diet-related risk factor for hypertension; thus, reducing dietary sodium has been identified as a cornerstone in medical nutritional therapy for managing hypertension and other CVDs [2]. The World Health Organization (WHO) recommends a daily sodium intake of up to 2000 mg [3]. However, the adult Mexican population has been reported to consume 1.5 times this amount, averaging between 3100 and 3500 mg per day [1,2,4]. Notably, high sodium intake is often accompanied by low potassium intake [1,2,4], which is considered an indicator of poor diet quality [5]. Additionally, the combination of high sodium and low potassium intake is linked to an increased cardiovascular risk in a dose-response relationship [6].

Several policies and campaigns have been implemented to support sodium reduction at the population level, including labeling products high in sodium for easier identification and removing table salt from restaurant tables. Preventable risk models indicate that reducing salt/sodium consumption could prevent a significant number of deaths from CVDs [7].

By contrast, patients' beliefs and perceived knowledge are strong predictors of their attitudes (ie, actions) and may, therefore, directly impact adherence to medical nutrition therapy [8,9]. Knowledge is defined as the understanding of a subject or topic, including the cognitive ability to retain such information [10]. Attitudes refer to the emotional, motivational, perceptual, and cognitive beliefs that positively or negatively influence a person's behavior. They affect future behavior independently of an individual's knowledge, helping to explain, at least in part, why a person adopts one behavior over others [10]. Finally, behaviors are defined as a set of responses of an individual or organism to external stimuli or internal motivation [11].

Although people may be aware of the health implications of high salt or sodium intake, their knowledge, attitudes, and behaviors (KAB) on this issue can still be lacking, as shown in a systematic review of 24 studies across 12 high-income countries [12]. Moreover, studies on the association between KAB and actual sodium intake have yielded mixed results in diverse populations; some studies report no association between salt- or sodium-related KAB and sodium excretion, while others have found certain associations [13-20]. Additionally, the relationship between KAB and potassium intake has received less attention [15]. In Mexico, salt- and sodium-related KAB and its association with sodium and potassium intake have not yet been studied.

Given this context, we aimed to describe self-reported KAB related to salt and sodium consumption in a sample of volunteers from Mexico City and the metropolitan area. Additionally, we explored the association between salt- and sodium-related KAB and 24-hour urinary sodium and potassium excretion as a surrogate for actual intake.

## Methods

### Study Design

This is a secondary analysis of data from a cross-sectional study designed to test the criterion validity of an online tool [21] for estimating sodium intake in the adult population. The validation results are currently under review for publication elsewhere.

### Study Population and Setting

The original study included men and women over 18 years of age who were apparently healthy. Exclusion criteria were individuals over 65 years; those with a previous diagnosis of CVD, kidney disease, or liver disease; and those who were menstruating, pregnant, or had urinary tract infections at the time of 24-hour urine collection. A convenience sample of participants meeting the eligibility criteria was recruited from various workplaces in Mexico City and the metropolitan area from June 2022 to January 2023. Participants were invited to join through printed advertisements on bulletin boards and a word-of-mouth strategy. Once eligibility was confirmed, they received both verbal and written instructions to collect a 24-hour urine sample and maintain a 3-day food record, which they were to bring on the day of their study visit. All assessments took place at designated areas within each work center.

For this secondary analysis, participants who provided a complete 24-hour urine sample and completed the salt/sodium-related KAB survey were included.

### Ethical Considerations

This study posed no risk to participants, and their data were handled with privacy and confidentiality. Participants' data were deidentified for analysis, with only the principal investigator having access to the identified data. The study was

approved by the Research and Ethics Board of the Instituto Nacional de Ciencias Médicas y Nutrición, under registration number 3314. All participants were fully informed about the study, and written informed consent was obtained. No compensation was offered to participants for their involvement in this study, other than the provision of the results from the body composition assessment conducted as part of the original study, as well as the results of the urinary analysis.

## Assessments

### *Anthropometrics*

Weight and height were measured with the participant fasting, without shoes, and wearing a hospital gown, following the guidelines established by the International Society for the Advancement of Kinanthropometry (ISAK) [22]. Weight was measured using a Seca 769 mechanical column scale with a capacity of 200 kg and a precision of 0.05 kg. Height was measured with a Seca 220 stadiometer with a precision of 1 mm. BMI was calculated and reported as $kg/m^2$.

### *24-Hour Urinary Sodium and Potassium*

Participants were given both oral and written instructions to collect a 24-hour urine sample the day before the study visit. They were instructed to discard the first-morning void and collect all urine over the following 24 hours, including the first void of the next morning, using a preservative-free container. Participants were also instructed to store the collected urine in a cool place during the collection period. Twenty-four-hour urinary excretion of sodium, potassium, and creatinine was determined.

Urinary samples were considered complete if they met the creatinine excretion rate criteria: 15-25 mg/kg/24 hours for men and 10-20 mg/kg/24 hours for women [23]. Additionally, urine samples with a volume of less than 500 ml per 2 hours were considered incomplete.

### *Energy Intake*

Participants were instructed to complete a 3-day food record, including 1 weekend day. On the day of the study visit, a member of the research team verified the completeness and clarity of the food record with the participant. Dietary information was collected by trained personnel using a nutrient analysis software program (ESHA Food Processor SQL version 10.11; ESHA Research). The mean energy intake from the 3 days was used for analysis.

### *Knowledge, Attitudes, and Behaviors*

A self-administered questionnaire developed by a subgroup of the Pan American Health Organization (PAHO) expert group, tasked with examining excessive dietary salt as a health risk in the Americas, was used. This tool was designed to assess salt-related knowledge (eg, the difference between sodium and salt), attitudes (eg, concern about reducing salt consumption), and behavior (eg, frequency of adding salt to food during cooking and at the table), self-reported presence of chronic diseases, and labeling preferences. The questionnaire was field-tested in Latin America and Canada. The questions included in this tool were developed based on the experience and expertise of the expert group members, as well as questions used in other surveys [24].

Participants received guidance from research team members to clarify any unclear questions as needed to complete the questionnaire.

Questions regarding labeling preferences (eg, whether participants would like food labeling indicating high/medium/low levels of salt or sodium) were not included in the analysis, as they were beyond the scope of this study.

### *Other Data*

Information on age, sex, and highest level of education was also collected.

## Statistical Analysis

Descriptive statistics were stratified by sex and reported as mean (SD) or median (25th-75th percentiles) for continuous variables, based on their distribution as determined by the Kolmogorov-Smirnov test. Categorical variables were presented as absolute frequencies and proportions. Comparisons of KAB related to health and salt/sodium consumption categories between sex groups were made using the chi-square test. In addition, one-way fixed-effects analyses of covariance were used to compare the adjusted means of 24-hour urinary sodium and potassium excretion across the KAB related to salt/sodium consumption categories (groups) for each KAB item. Separate models were constructed for 24-hour urinary sodium and potassium excretion. In these models, urinary sodium and potassium excretion were treated as the dependent variables, with each relevant item of salt/sodium-related KAB tested separately as the independent variable (factor) with multiple categories (groups). Factors known to be associated with sodium and potassium intake (age, sex, BMI, and daily energy intake [DEI]) [2,4] were included as covariates in all models. The general linear model used to compare 24-hour urinary sodium and potassium excretion among salt/sodium-related KAB groups, adjusting for the effect of covariates, can be expressed as follows [25,26]:

$$\text{24-hour UNa}_{ij} = \mu + \alpha_i + \beta_1 \text{age}_{ij} + \beta_2 \text{sex}_{ij} + \beta_3 \text{BMI}_{ij} + \beta_4 \text{DEI}_{ij} + e_{ij}$$

$$\text{24-hour UK}_{ij} = \mu + \alpha_i + \beta_1 \text{age}_{ij} + \beta_2 \text{sex}_{ij} + \beta_3 \text{BMI}_{ij} + \beta_4 \text{DEI}_{ij} + e_{ij}$$

where 24-h $UNa_{ij}$ is the 24-hour urinary sodium excretion of the *j*th participant in the *i*th salt/sodium-related KAB group, 24-h $UK_{ij}$ is the 24-hour urinary potassium excretion of the *j*th participant in the *i*th salt/sodium-related KAB group, μ is a grand mean, $a_i$ is the *i*th group effect, $\beta_1$ is the slope coefficient of age, $\beta_2$ is the slope coefficient of sex, $\beta_3$ is the slope coefficient of BMI, $\beta_4$ is the slope coefficient of DEI, and $e_{ij}$ is the residual.

95% CIs were constructed for the adjusted means, and the Šidák correction was applied for multiple comparisons. All statistical analyses were performed using SPSS version 20 for Windows (IBM Corp.). A *P* value of <.05 was considered statistically significant.

# Results

## Study Sample Descriptive Characteristics

Overall, 365 participants were recruited. Of these, 232 completed the salt/sodium-related KAB questionnaire and provided a valid 24-hour urinary sodium sample; these participants were included in this subanalysis. Study sample characteristics by sex are shown in Table 1. The majority of the study sample were women (184/232, 79.3%), with a median age of 39.0 (25th-75th percentiles 27.3-49.0) years and a BMI of 25.8 (25th-75th percentiles 22.9-29.2) kg/m$^2$ for the overall population. In terms of education, most participants had completed an education level higher than secondary (101/232, 43.5%). The median energy intake was estimated to be 2704.7 (25th-75th percentiles 2128.0-3271.1) kcal/day. Estimates of urinary sodium, potassium, and creatinine are also shown in Table 1.

**Table 1.** Study sample characteristics by sex[a].

| Characteristics | Overall (n=232) | Women (n=184) | Men (n=48) |
|---|---|---|---|
| Age (years) | 39.0 (27.3-49.0) | 41 (29-50) | 36.9 (12.5) |
| Weight (kg) | 63.7 (55.5-71.8) | 62.7 (12.3) | 70.8 (62.4-80.4) |
| Height (cm) | 156.0 (150-162.8) | 154.0 (6.8) | 167.1 (8.2) |
| BMI (kg/m$^2$) | 25.8 (22.9-29.2) | 26.4 (5.0) | 25.6 (23.6-27.4) |
| **Education level** | | | |
| None | 6 (2.6) | 6 (3.3) | 0 (0) |
| Primary | 29 (12.5) | 25 (13.6) | 4 (8.3) |
| Secondary | 93 (40.1) | 72 (39.1) | 21 (43.8) |
| Higher | 101 (43.5) | 79 (42.9) | 22 (45.8) |
| No answer | 3 (1.3) | 2 (1.1) | 1 (2.1) |
| Energy intake (kcal/day) | 2704.7 (2128.0-3271.1) | 2726.6 (2141.5-3254.4) | 2778.2 (981.8) |
| Diuresis (ml/24 hours) | 1225.0 (900.0-1700.0) | 1200 (900-1587.5) | 1400.0 (1000.0-2000.0) |
| Sodium (mg/24 hours) | 2582.5 (1701.6-3347.2) | 2570.4 (1222.3) | 2845.1 (2231.4-3631.1) |
| Potassium (mg/24 hours) | 1493.5 (1096.3-1934.7) | 1431.9 (1071.0-1894.5) | 1873.2 (736.4) |
| Sodium-to-potassium ratio (mg/mg/24 hours) | 1.7 (1.3-2.3) | 1.7 (1.3-2.3) | 1.6 (1.3-2.1) |
| Creatinine (mg/24 hours) | 87.1 (61.4-119.7) | 81.0 (56.6-111.11) | 108.8 (86.2-163.2) |

[a]Numerical data are presented as mean (SD) or median (25th-75th percentiles), according to their distribution. Categorical data are presented as n (%).

Regarding the self-reported presence of chronic diseases included in the survey, 43 of the 232 (18.5%) participants reported having high blood pressure. Other pathologies associated with excessive salt consumption were also surveyed and are reported, stratified by sex, in Table 2.

**Table 2.** Self-reported history of chronic diseases by sex.

| Disease | Overall (n=232) | Women (n=184) | Men (n=48) |
|---|---|---|---|
| High blood pressure, n (%) | 43 (18.5) | 40 (21.7) | 3 (6.3) |
| Heart attack, n (%) | 7 (3.0) | 6 (3.3) | 1 (2.1) |
| Stroke, n (%) | 1 (0.4) | 1 (0.5) | 0 (0) |
| Kidney stones, n (%) | 14 (6.0) | 10 (5.4) | 4 (8.3) |
| Asthma, n (%) | 10 (4.3) | 10 (5.4) | 0 (0) |
| Osteoporosis, n (%) | 2 (0.9) | 2 (1.1) | 0 (0) |
| Stomach cancer, n (%) | 1 (0.4) | 1 (0.5) | 0 (0) |

## Attitudes, Knowledge, and Behaviors

As shown in Table 3, more than half of the participants reported trying to eat a healthy diet and consistently feeling pressure to do so. Nearly the entire sample (229/232, 98.7%) acknowledged that excessive salt is associated with health problems, although only two-thirds believed they were capable of distinguishing between products low or high in salt. Furthermore, less than half of the sample indicated that there was sufficient information on salt content in food packages. Notably, only 93 of the 232 (40.1%) participants reported overall good health, with a trend toward a lower proportion of women compared with men

(71/184, 38.6%, vs 22/48, 45.8%, *P*=.07). The number of participants who reported trying to minimize the amount of salt they consumed was 174 (75%) in the overall population (N=232), and this tendency was higher among women than men (143/184, 77.7%, vs 31/48, 64.6%, *P*=.08). However, no significant differences were found between men and women in any of the questions (*P*≥.05; Table 3).

**Table 3.** The proportion of participants who agreed with the statement on self-reported attitudes, knowledge, and behavior related to health, diet, and salt intake in the study sample.

| Statement | Overall (n=232) | Women (n=184) | Men (n=48) | *P* value[a] |
|---|---|---|---|---|
| I try to eat a healthy diet, n (%) | 153 (65.9) | 120 (65.2) | 33 (68.8) | .71 |
| A diet high in salt can cause serious health problems, n (%) | 229 (98.7) | 182 (98.9) | 47 (97.9) | .52 |
| I try to minimize the amount of fat I consume, n (%) | 187 (80.6) | 149 (81.0) | 38 (79.2) | .68 |
| My health is generally good, n (%) | 93 (40.1) | 71 (38.6) | 22 (45.8) | .07 |
| There is too much pressure to eat healthy these days, n (%) | 134 (57.8) | 109 (59.2) | 25 (52.1) | .76 |
| I try to minimize the amount of salt I consume, n (%) | 174 (75.0) | 143 (77.7) | 31 (64.6) | .08 |
| I generally know if foods contain a lot or little salt, n (%) | 148 (63.8) | 122 (66.3) | 26 (54.2) | .12 |
| There is sufficient nutritional information on food packaging, n (%) | 106 (45.7) | 84 (45.7) | 22 (45.8) | .34 |

[a]For differences between sex groups by chi-square test.

Additionally, 150 of the 232 (64.7%) participants reported always adding salt when preparing food at home, while more than 90% considered limiting the amount of salt or sodium in their diet to be very or somewhat important (210/232, 90.5%). Around 60% believed they consumed the right amount of salt (139/232, 59.9%), yet less than a quarter knew there was a recommended daily amount of sodium consumption; 186 of 232 (80.2%) participants did not know the difference between salt and sodium, and over half reported rarely or never reading nutrition labels on food packages. Comparison between sex groups revealed a higher proportion of men who did not know the amount of sodium they consumed compared with women (12/48, 25%, vs 15/184, 8.2%, *P*=.01), while more women reported knowing that there was a recommended daily sodium intake than men (46/184, 25%, vs 10/48, 20.8%, *P*=.02). Notably, there was a difference in the frequency of reading nutrition labels on food packages between sex groups, with more than half of men (30/48, 62.5%) reporting that they never or rarely read food labels, compared with women (96/184, 52.1%, *P*=.04; Table 4).

**Table 4.** Answers to questions about knowledge, attitudes, and behaviors regarding sodium intake.

| Question | Overall (n=232) | Women (n=184) | Men (n=48) | *P* value[a] |
|---|---|---|---|---|
| **How many times do you add salt to food at the table?, n (%)** | | | | .12 |
| Never | 40 (17.2) | 32 (17.4) | 8 (16.7) | |
| Rarely | 102 (44.0) | 85 (46.2) | 17 (35.4) | |
| Sometimes | 58 (25.0) | 45 (24.5) | 13 (27.1) | |
| Often | 20 (8.6) | 16 (8.7) | 4 (8.3) | |
| Always | 12 (5.2) | 6 (3.3) | 6 (12.5) | |
| **Do you add salt when preparing food at home?, n (%)** | | | | .12 |
| Never | 7 (3.0) | 3 (1.6) | 4 (8.3) | |
| Rarely | 12 (5.2) | 10 (5.4) | 2 (4.2) | |
| Sometimes | 24 (10.3) | 17 (9.2) | 7 (14.6) | |
| Often | 39 (16.8) | 32 (17.4) | 7 (14.6) | |
| Always | 150 (64.7) | 122 (66.3) | 28 (58.3) | |
| **Is limiting the amount of salt or sodium in your diet important to you?, n (%)** | | | | .23 |
| Very | 95 (40.9) | 77 (41.8) | 18 (37.5) | |
| Somewhat | 115 (49.6) | 93 (50.5) | 22 (45.8) | |
| No | 14 (6.0) | 8 (4.3) | 6 (12.5) | |
| Do not know | 6 (2.6) | 4 (2.2) | 2 (4.2) | |
| No answer | 2 (0.9) | 2 (1.1) | 0 (0) | |
| **How much salt do you think you consume?, n (%)** | | | | .01 |
| Not enough | 18 (7.8) | 16 (8.7) | 2 (4.2) | |
| Right amount | 139 (59.9) | 111 (60.3) | 28 (58.3) | |
| Too much | 46 (19.8) | 40 (21.7) | 6 (12.5) | |
| Do not know | 27 (11.6) | 15 (8.2) | 12 (25.0) | |
| No answer | 2 (0.9) | 2 (1.1) | 0 (0) | |
| **Do you know if there is a recommended amount of salt/sodium consumption per person per day?, n (%)** | | | | .02 |
| Yes | 56 (24.1) | 46 (25.0) | 10 (20.8) | |
| No | 174 (75.0) | 138 (75.0) | 36 (75.0) | |
| No answer | 2 (0.9) | 0 (0) | 2 (4.2) | |
| **Do you know the difference between salt and sodium?, n (%)** | | | | .42 |
| Yes | 44 (19.0) | 33 (17.9) | 11 (22.9) | |
| No | 186 (80.2) | 150 (81.5) | 36 (75.0) | |
| No answer | 2 (0.9) | 1 (0.5) | 1 (2.1) | |
| **Do you pay attention to the text on the packaging such as no added salt, low salt, light, trans-fat free?, n (%)** | | | | .23 |
| Always | 12 (5.2) | 9 (4.9) | 3 (6.3) | |
| Often | 35 (15.1) | 31 (16.8) | 4 (8.3) | |
| Sometimes | 63 (27.2) | 53 (28.8) | 10 (20.8) | |
| Rarely | 67 (28.9) | 53 (28.8) | 16 (33.3) | |
| Never | 49 (21.1) | 37 (20.1) | 12 (25.0) | |
| Do not know | 6 (2.6) | 3 (1.6) | 3 (6.3) | |
| **How often do you read nutrition labels on food packages?, n (%)** | | | | .04 |

| Question | Overall (n=232) | Women (n=184) | Men (n=48) | *P* value[a] |
|---|---|---|---|---|
| Always | 12 (5.2) | 8 (4.3) | 4 (8.3) | |
| Often | 26 (11.2) | 20 (10.9) | 6 (12.5) | |
| Sometimes | 67 (28.9) | 60 (32.6) | 7 (14.6) | |
| Rarely | 91 (39.2) | 72 (39.1) | 19 (39.6) | |
| Never | 35 (15.1) | 24 (13.0) | 11 (22.9) | |
| No answer | 1 (0.4) | 0 (0) | 1 (2.1) | |

[a]For differences between sex groups by chi-square test.

### Salt/Sodium-Related Attitudes, Knowledge, and Behaviors and 24-Hour Urinary Sodium and Potassium Excretion

A comparison of the adjusted means (95% CI) for 24-hour urinary sodium and potassium excretion across categories of salt/sodium-related KAB is shown in Table 5. A higher mean 24-hour urinary sodium excretion was observed among participants who reported consuming too much sodium, compared with those who believed they consumed the right amount. It was also higher among participants who knew that there was a recommended daily sodium intake and who understood the difference between salt and sodium, compared with those who did not. Additionally, participants who frequently paid attention to the text on food packaging had a higher mean urinary sodium excretion than those who rarely did so. Although the overall comparison for the frequency of reading nutrition labels showed a significant difference among frequency categories ($P$=.02), pairwise comparisons did not reveal any significant differences ($P$≥.05 for all pairwise comparisons).

Regarding urinary potassium excretion, participants who tried to eat a healthy diet had a higher mean potassium excretion compared with those who were unsure if they were trying to eat a healthy diet. Additionally, although a significant difference in potassium excretion was observed in the overall comparison among categories of the amount of salt participants thought they consumed ($P$=.02), pairwise comparisons did not reveal any significant differences ($P$≥.05 for all pairwise comparisons). Higher potassium excretion was also observed among participants who knew there was a recommended daily sodium intake and those who claimed to know the difference between salt and sodium, compared with those who did not. In terms of reading food labels, participants who reported often paying attention to the text on food packaging had greater potassium excretion compared with those who rarely did so. Similarly, individuals who always read nutrition food labels had higher potassium excretion compared with those who sometimes, rarely, or never read food labels (Table 5).

**Table 5.** Comparison of urinary sodium and potassium excretion adjusted means (95% CI) among groups within the items of knowledge, attitudes, and behaviors regarding sodium intake.

| Question | 24-h urinary sodium excretion (mg/day), adjusted means (95% CI) | 24-h urinary potassium excretion (mg/day), adjusted means (95% CI) |
|---|---|---|
| **I try to eat a healthy diet** | | |
| Agree (n=153) | 2802.9 (2606.3-2999.6) | 1666.5 (1561.1-1771.9)[a] |
| Disagree (n=31) | 2396.6 (1956.7-2836.4) | 1557.2 (1321.4-1793.0) |
| Do not know (n=35) | 2576.2 (2164.1-2988.3) | 1331.5 (1110.5-1552.4)[a] |
| *P* value | .21 | .03 |
| **A diet high in salt can cause serious health problems** | | |
| Agree (n=229) | 2696.7 (2538.1-2855.4) | 1594.6 (1508.6-1680.5) |
| Do not know (n=2) | 1900.0 (195.2-3604.7) | 1530.6 (607.2-2454.0) |
| *P* value | .36 | .89 |
| **I try to minimize the amount of salt I consume** | | |
| Yes (n=174) | 2732.0 (2549.9-2914.2) | 1603.7 (1504.3-1703.2) |
| No (n=40) | 2416.5 (2035.4-2797.6) | 1528.7 (1320.6-1736.8) |
| Do not know (n=13) | 2598.1 (1927.7-3268.4) | 1631.8 (1265.7-1997.8) |
| *P* value | .34 | .80 |
| **I generally know if foods contain a lot or little salt** | | |
| Yes (n=148) | 2744.3 (2545.4-2943.2) | 1636.3 (1528.7-1743.8) |
| No (n=80) | 2608.4 (2336.8-2880.1) | 1517.4 (1370.5-1664.3) |
| *P* value | .43 | .20 |
| **How many times do you add salt to food at the table?** | | |
| Never (n=40) | 2766.4 (2384.4-3148.5) | 1624.1 (1416.5-1831.7) |
| Rarely (n=102) | 2686.9 (2446.9-2926.9) | 1572.0 (1441.5-1702.4) |
| Sometimes (n=58) | 2638.2 (2320.4-2956.0) | 1570.5 (1397.9-1743.2) |
| Often (n=20) | 2897.6 (2354.7-3440.6) | 1630.4 (1335.4-1925.5) |
| Always (n=12) | 2221.2 (1514.6-2927.8) | 1663.1 (1279.2-2047.1) |
| *P* value | .64 | .98 |
| **Do you add salt when preparing food at home?** | | |
| Never (n=7) | 2561.7 (1637.7-3485.6) | 1943.3 (1437.9-2448.7) |
| Rarely (n=12) | 3656.2 (2966.3-4346.0) | 1748.9 (1371.6-2126.3) |
| Sometimes (n=24) | 2694.3 (2206.9-3181.7) | 1478.5 (1211.9-1745.1) |
| Often (n=39) | 2709.1 (2321.9-3096.3) | 1586.4 (1374.6-1798.2) |
| Always (n=150) | 2601.5 (2405.8-2797.1) | 1580.1 (1473.1-1687.1) |
| *P* value | .08 | .51 |
| **How much salt do you think you consume?** | | |
| Not enough (n=18) | 2346.7 (1788.3-2905.0) | 1396.8 (1090.9-1702.8) |
| Right amount (n=139) | 2584.3 (2384.9-2783.7)[b] | 1550.0 (1440.7-1659.2) |
| Too much (n=46) | 3216.0 (2867.1-3565.0)[b] | 1845.3 (1654.1-2036.5) |
| Do not know (n=27) | 2471.7 (2000.1-2943.2) | 1484.5 (1226.2-1742.9) |
| *P* value | .008 | .02 |
| **Is limiting the amount of salt or sodium in your diet important to you?** | | |

| Question | 24-h urinary sodium excretion (mg/day), adjusted means (95% CI) | 24-h urinary potassium excretion (mg/day), adjusted means (95% CI) |
|---|---|---|
| Very (n=95) | 2800.6 (255.8-3050.3) | 1558.6 (1423.7-1693.4) |
| Somewhat (n=115) | 2630.5 (2405.1-2855.9) | 1632.0 (1510.3-1753.7) |
| No (n=14) | 2373.4 (1712.3-3034.5) | 1525.4 (1168.5-1882.3) |
| Do not know (n=6) | 3005.0 (2017-3993.0) | 1756.7 (1223.4-2290.1) |
| *P* value | .52 | .77 |
| **Do you know if there is a recommended amount of salt/sodium consumption per person per day?** | | |
| Yes (n=56) | 3056.7 (2733.1-3380.2) | 1882.5 (1711.4-2053.6) |
| No (n=174) | 2562.9 (2381.5-2744.2) | 1488.3 (1392.3-1584.2) |
| *P* value | .01 | <.001 |
| **Do you know the difference between salt and sodium?** | | |
| Yes (n=44) | 3011.5 (2640.1-3382.9) | 1864.9 (1669.6-2060.3) |
| No (n=186) | 2592.8 (2417.2-2768.3) | 1512.5 (1420.1-1604.8) |
| *P* value | .049 | .002 |
| **Do you pay attention to the text on the packaging such as no added salt, low salt, light, trans-fat free?** | | |
| Always (n=12) | 3034.4 (2348.8-3720.1) | 1673.2 (1305.1-2041.4) |
| Often (n=35) | 3278.0 (2870.2-3685.8)[c] | 1859.5 (1640.6-2078.5)[d] |
| Sometimes (n=63) | 2571.7 (2270.4-2873.0) | 1692.0 (1530.2-1853.7) |
| Rarely (n=67) | 2501.3 (2210.5-2792.1)[c] | 1390.7 (1234.5-1546.8)[d] |
| Never (n=49) | 2607.0 (2257.2-2956.7) | 1559.3 (1371.5-1747.1) |
| Do not know (n=6) | 2308.9 (1332.1-3285.6) | 1269.7 (745.3-1794.2) |
| *P* value | .04 | .01 |
| **How often do you read nutrition labels on food packages?** | | |
| Always (n=12) | 3456.4 (2764.0-4148.9) | 2158.3 (1789.7-2526.9)[e,f,g] |
| Often (n=26) | 3214.7 (2739.9-3689.4) | 1921.1 (1668.4-2173.8)[h] |
| Sometimes (n=67) | 2608.5 (2313.5-2903.6) | 1560.7 (1403.6-1717.7)[e] |
| Rarely (n=91) | 2519.5 (2270.8-2768.1) | 1473.4 (1341.0-1605.7)[f,h] |
| Never (n=35) | 2583.2 (2170.2-2996.3) | 1509.9 (1290.0-1729.8)[g] |
| *P* value | .02 | .001 |

[a]*P*=.02; [b]*P*=.01; [c]*P*=.04; [d]*P*=.01; [e]*P*=.04; [f]*P*=.007; [g]*P*=.03; [h]*P*=.02 for multiple comparisons by the Šidák test. Differences are between categories that have the same letter. All means were adjusted by sex, age, BMI, and daily energy intake by analysis of covariance.

## Discussion

### Principal Findings and Comparison With Prior Work

Based on the data collected and analyzed in this study, several important findings can be reported regarding consumers' attitudes, knowledge, and behavior concerning salt intake, and their association with urinary sodium and potassium excretion.

According to the self-reported KAB related to health, diet, and salt consumption, among the total participants (N=232), only 56 (24.1%) knew there was a recommended daily intake for salt/sodium, 44 (19%) understood the difference between salt and sodium, and only 12 (5.2%) always paid attention to the text on food packaging or read nutrition labels. Importantly, a high proportion of participants reported trying to maintain a healthy diet (153/232, 65.9%) and reduce the amount of fats (187/232, 80.6%) or salt (174/232, 75%) they consumed. However, these responses did not align with their perception of health, as only 93 (40.1%) considered their health to be generally good. It is likely that participants were making these dietary changes in an effort to improve their health, which they perceived as poor.

In our study, we observed a trend where a smaller proportion of women (71/184, 38.6%) compared with men (22/48, 45.8%) reported having good health. Additionally, self-reported history of comorbidities related to excessive sodium intake revealed that 40 of 184 (21.7%) women and 3 of 48 (6.3%) men had high blood pressure. This lower perception of good health and higher frequency of self-reported high blood pressure among women is consistent with the observed trend of a higher proportion of women trying to minimize their salt consumption. Overall, studies that disaggregate KAB by sex have reported better KAB related to salt/sodium consumption in women than in men. In this study, we found that more women (46/184, 25.0%) reported knowing that there was a recommended daily sodium/salt intake compared with men (10/48, 20.8%). Additionally, more men (30/48, 62.5%) reported never or rarely reading food labels compared with women (96/184, 52.2%).

In the overall study sample, the median 24-hour urinary sodium excretion was 2582.5 mg/day, which exceeds the WHO daily intake recommendation of 2000 mg/day [3], but is lower than previous reports in the Mexican population, where it ranged from 3100 to 3500 mg/day [1,2,4]. This discrepancy may be attributed to the fact that nearly 80% (n=184) of the participants in this study were women, and sodium intake has been shown to be, on average, lower in women than in men [2,4], as also confirmed in this study; 32 out of 232 (13.8%) reported always or often adding salt to food at the table; however, a higher proportion (189/232, 81.5%) of participants reported always or often adding salt during food preparation. In Mexico, the consumption of processed and ultra-processed foods contributes to 39% of total sodium intake in adults and 50% in school-age children [27]. Nevertheless, salt added during cooking remains an important source of sodium in the diet [27]. In fact, it has been reported that street food (*antojitos* in Spanish) and homemade-style meals (*comida corrida* in Spanish), which are unpackaged and unlabeled prepared foods, contain high levels of sodium, with many providing more than 25% of the daily sodium intake recommended by the WHO [28]. Thus, the use of table salt in food preparation remains a significant challenge in Mexico, as it does in other countries such as Brazil, China, Costa Rica, Guatemala, India, Japan, Mozambique, and Romania, where a large proportion of overall daily sodium intake comes from the discretionary use of salt [29].

In terms of potassium intake, the median urinary excretion was 1493.5 mg/day in the overall study sample, which is considerably lower than the WHO's recommended intake of 3510 mg/day for adults to prevent chronic disease and improve health [30]. Low urinary potassium excretion was also reported in 2 studies of the Mexican population, with values ranging from 1909 to 1981 mg/day [2,4], while another report indicated a higher mean potassium intake (3401 mg/day) in this population, as determined by a 24-hour food recall [27]. Importantly, potassium intake, whether determined by urinary excretion or dietary analysis, has been found to be an indicator of diet quality [5,31] and is inversely associated with BMI, diastolic blood pressure, and heart rate [31]. Additionally, the combination of higher sodium and lower potassium intakes, as observed in this and other studies [2,4,27], is associated with an increased cardiovascular risk [6]. Therefore, dietary strategies aimed at modifying both sodium and potassium intakes are necessary for the Mexican population.

Diverse studies have reported no association between salt/sodium-related KAB [13,17,18,20] and sodium excretion, while others have found some level of association [14,15,19]. To explore the relationship between salt/sodium-related KAB and sodium and potassium intake, we compared the adjusted means of 24-hour urinary sodium and potassium excretion across categories of salt/sodium-related KAB. Interestingly, the mean sodium excretion was similar among participants, regardless of whether they were trying to eat a healthy diet. However, potassium excretion was higher among participants who reported trying to eat a healthy diet compared with those who were unsure, after adjusting for sex, age, BMI, and calorie intake. Similarly, participants who often paid attention to the text on packaging had higher levels of both sodium and potassium excretion compared with those who rarely did so. Potassium excretion was also higher in participants who always read nutrition food labels compared with those who did so less frequently, although this behavior was not associated with sodium excretion in the multiple comparisons. These results suggest that attempts to follow a healthy diet may impact the quality of the diet by likely incorporating more food sources of potassium, such as fruits, vegetables, and legumes, but not necessarily modifying the selection of foods based on their sodium content. Another possibility is that to reduce sodium intake, people may increase the consumption of low-sodium processed foods that contain hidden potassium in the form of additives, such as sodium-reduced meat and poultry products and industrialized bakery items [32,33]. Further research is needed to clarify these findings.

Knowing that there is a recommended amount of daily salt/sodium consumption or understanding the differences between salt and sodium was not associated with lower sodium intake either; in fact, participants who reported knowing this exhibited higher sodium and potassium excretion than those who did not. These results are relevant as they support the findings of a study in a Chinese population, where self-reported salt/sodium-related KAB had a smaller effect on sodium but a greater effect on potassium excretion [15].

The results of this study provide evidence for the need to continue improving KAB related to salt/sodium consumption at the population level. The implementation of education and communication campaigns is one of the strategies endorsed by the WHO to raise awareness about the health risks and dietary sources of salt, aiming to induce behavioral changes [34]. In Mexico, the "Less Salt, More Health" campaign was implemented in 2013 by the Ministry of Health. This communication strategy aimed to increase knowledge and educate the population about the benefits of reducing salt consumption. However, the impact of its implementation was not assessed [1]. In fact, this is the first study to report KAB related to salt/sodium intake in a Mexican population.

Worldwide, other strategies beyond communication campaigns that have shown the greatest impact on dietary sodium reduction at the population level include mandatory food reformulation, food labeling, and taxes [35]. In 2020, Mexico implemented a

mandatory front-of-pack warning labeling system [36]. The reformulation of processed and ultra-processed foods that contain large amounts of sodium should be one of the next steps in Mexico's sodium reduction strategy. While some work has been done in this area, further progress is needed in line with the WHO regional sodium benchmarks [1].

Finally, even though better salt/sodium-related KAB was not associated with lower sodium excretion, participants had some awareness of their actual sodium intake level. Specifically, those who reported consuming the right amount of dietary sodium had lower sodium excretion compared with those who believed they consumed too much salt. Nonetheless, participants who reported an adequate intake (the right amount) still had sodium consumption levels above the recommended 2000 mg/day. This underscores the ongoing need to improve salt/sodium-related KAB and further reduce sodium intake, as previously discussed.

### Limitations

This study has several limitations. First, its cross-sectional design prevents the establishment of causal relationships between salt/sodium-related KAB and urinary sodium and potassium excretion. Second, the study was conducted with a sample from Mexico City and its surrounding area, limiting the generalizability of the results to the national population. Third, nearly 80% (n=184) of the study sample were women; however, data were stratified by sex, and the adjusted urinary sodium and potassium means accounted for sex as a covariate. Fourth, the available data do not distinguish between potassium naturally occurring in foods and that added as an additive in processed foods. Further studies are needed to explore the sources of dietary potassium in the Mexican population and its impact on health-related outcomes. Finally, no sample size calculation was conducted; instead, all available data from participants meeting the eligibility criteria for this secondary analysis were analyzed. Longitudinal and properly powered studies exploring this association are warranted. The main strength of this study lies in the use of an objective method to assess sodium and potassium intake, namely, the measurement of their excretion levels in 24-hour urine.

### Conclusions

Results of this study indicate that self-reported KAB related to salt/sodium intake is poor in this study sample. Therefore, new initiatives are needed to improve KAB in order to promote better food choices, which, in turn, could help reduce sodium intake. In this context, educating individuals about the sodium content of unpackaged prepared foods is particularly relevant. In addition, salt/sodium-related KAB had a greater impact on potassium excretion than on sodium excretion, highlighting the need for more strategies to improve KAB related to salt/sodium intake. It also emphasizes the importance of considering other approaches to modify the sodium content of foods, such as the reformulation of processed and ultra-processed foods. Finally, monitoring KAB related to salt/sodium consumption, along with actual sodium intake, in larger populations is essential, especially when new strategies to reduce sodium intake are implemented.

## Data Availability

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

## Authors' Contributions

GGS and PVMA contributed to the conceptualization, methodology, and writing of the original draft. PBA, LEOR, NIV, AEC, and RCR were involved in the investigation, reviewing, and editing of the original draft; ECR contributed to conceptualization, methodology, formal analysis, project administration, reviewing, and editing of the original draft. All authors critically revised the manuscript and approved the final version.

## Conflicts of Interest

None declared.

## References

1. Campos-Nonato I, Vargas Meza J, Nieto C, Ariza AC, Barquera S. Reducing sodium consumption in Mexico: a strategy to decrease the morbidity and mortality of cardiovascular diseases. Front Public Health 2022;10:857818 [FREE Full text] [doi: 10.3389/fpubh.2022.857818] [Medline: 35392467]
2. Vega-Vega O, Fonseca-Correa J, Mendoza-De la Garza A, Rincón-Pedrero R, Espinosa-Cuevas A, Baeza-Arias Y, et al. Contemporary dietary intake: too much sodium, not enough potassium, yet sufficient iodine: the SALMEX cohort results. Nutrients 2018 Jun 25;10(7):816 [FREE Full text] [doi: 10.3390/nu10070816] [Medline: 29941792]
3. World Health Organization (WHO). Guideline: Sodium Intake for Adults and Children. Geneva, Switzerland: World Health Organization (WHO); 2012.
4. Vallejo M, Colín-Ramírez E, Rivera Mancía S, Cartas Rosado R, Madero M, Infante Vázquez O, et al. Assessment of sodium and potassium intake by 24 h urinary excretion in a healthy Mexican cohort. Arch Med Res 2017 Feb;48(2):195-202. [doi: 10.1016/j.arcmed.2017.03.012] [Medline: 28625323]
5. Farapti F, Buanasita A, Atmaka D, Setyaningtyas S, Adriani M, Rejeki P, et al. Potassium intake is associated with nutritional quality and actual diet cost: a study at formulating a low sodium high potassium (LSHP) healthy diet. J Nutr Sci 2022;11:e11 [FREE Full text] [doi: 10.1017/jns.2021.104] [Medline: 35291270]

6. Ma Y, He FJ, Sun Q, Yuan C, Kieneker LM, Curhan GC, et al. 24-Hour Urinary Sodium and Potassium Excretion and Cardiovascular Risk. N Engl J Med 2022 Jan 20;386(3):252-263. [doi: 10.1056/nejmoa2109794]
7. Vargas-Meza J, Nilson EAF, Nieto C, Khandpur N, Denova-Gutiérrez E, Valero-Morales I, et al. Modelling the impact of sodium intake on cardiovascular disease mortality in Mexico. BMC Public Health 2023 May 26;23(1):983. [doi: 10.1186/s12889-023-15827-0]
8. Náfrádi L, Nakamoto K, Schulz PJ. Is patient empowerment the key to promote adherence? A systematic review of the relationship between self-efficacy, health locus of control and medication adherence. PLoS One 2017 Oct;12(10):e0186458 [FREE Full text] [doi: 10.1371/journal.pone.0186458] [Medline: 29040335]
9. West LM, Borg Theuma R, Cordina M. Health locus of control: Its relationship with medication adherence and medication wastage. Research in Social and Administrative Pharmacy 2018 Nov;14(11):1015-1019. [doi: 10.1016/j.sapharm.2017.12.003]
10. Fautsch MY, Glasauer P. Guidelines for assessing nutrition-related Knowledge, Attitudes and Practices. Roma: Food and Agriculture Organization of the United Nations; 2014.
11. Galimberti U. Diccionario de Psicología. México: Siglo Veintiuno editores; 2002.
12. Bhana N, Utter J, Eyles H. Knowledge, Attitudes and Behaviours Related to Dietary Salt Intake in High-Income Countries: a Systematic Review. Curr Nutr Rep 2018 Sep 3;7(4):183-197. [doi: 10.1007/s13668-018-0239-9]
13. Moliterno P, Donangelo C, Borgarello L, Oviedo L, Nogara R, Olascoaga A. Salt-related practices and its association with 24-hour urinary sodium excretion in an Uruguayan population cohort. Rev Panam Salud Publica 2022;46:e180. [doi: 10.26633/rpsp.2022.180]
14. Fan F, Li Y, Li L, Nie X, Zhang P, Li Y, et al. Salt-Related Knowledge, Attitudes, and Behaviors and Their Relationship with 24-Hour Urinary Sodium Excretion in Chinese Adults. Nutrients 2022 Oct 20;14(20):4404. [doi: 10.3390/nu14204404]
15. Du X, Fang L, Xu J, Chen X, Bai Y, Wu J, et al. The association of knowledge, attitudes and behaviors related to salt with 24-h urinary sodium, potassium excretion and hypertensive status. Sci Rep 2022 Aug 16;12(1):13901. [doi: 10.1038/s41598-022-18087-x]
16. Land M, Webster J, Christoforou A, Johnson C, Trevena H, Hodgins F, et al. The association of knowledge, attitudes and behaviours related to salt with 24-hour urinary sodium excretion. Int J Behav Nutr Phys Act 2014;11(1):47. [doi: 10.1186/1479-5868-11-47]
17. Marakis G, Marques Domingues A, Crispo A, Magriplis E, Vasara E, Kontopoulou L, et al. Pertinence of Salt-Related Knowledge and Reported Behaviour on Salt Intake in Adults: A Cross-Sectional Study. Nutrients 2023 Sep 23;15(19):4114. [doi: 10.3390/nu15194114]
18. Garg V, Shivashankar R, Kondal D, Ghosh S, Khandelwal S, Gupta R, et al. Knowledge, attitudes and practices related to dietary salt intake among adults in North India. Public Health Nutr 2019 Jun;22(09):1606-1614. [doi: 10.1017/s1368980018003518]
19. Johnson C, Mohan S, Rogers K, Shivashankar R, Thout S, Gupta P, et al. The Association of Knowledge and Behaviours Related to Salt with 24-h Urinary Salt Excretion in a Population from North and South India. Nutrients 2017 Feb 16;9(2):144. [doi: 10.3390/nu9020144]
20. Fernandez I, Godoy M, Feliu M, Vidueiros S, Pallaro A. Conocimientos, actitudes y comportamiento en torno a la sal alimentaria de adultos jóvenes universitarios: asociación con la ingesta de sal. Rev Esp Nutr Comunitaria 2018;24(3):1-12.
21. Colín-Ramírez E, Cartas-Rosado R, Vannesa P, Alatriste M, Cuevas, Arcand J. Development of an online tool for sodium intake assessment in Mexico. Rev Panam Salud Publica 2019 Dec 6;43:e90 [FREE Full text] [doi: 10.26633/rpsp.2019.90]
22. The International Society for the Advancement of Kinanthropometry (ISAK). International Standards for Anthropometric Assessment. Australia: ISAK; 2001.
23. Wielgosz A, Robinson C, Mao Y, Jiang Y, Campbell NRC, Muthuri S, et al. The Impact of Using Different Methods to Assess Completeness of 24‐Hour Urine Collection on Estimating Dietary Sodium. J of Clinical Hypertension 2016 Jun;18(6):581-584. [doi: 10.1111/jch.12716]
24. Claro RM, Linders H, Ricardo CZ, Legetic B, Campbell NRC. Consumer attitudes, knowledge, and behavior related to salt consumption in sentinel countries of the Americas. Rev Panam Salud Publica 2012 Oct;32(4):265-273. [doi: 10.1590/s1020-49892012001000004]
25. Shieh G. Power Analysis and Sample Size Planning in ANCOVA Designs. Psychometrika 2020 Mar;85(1):101-120. [doi: 10.1007/s11336-019-09692-3]
26. Rosner B. Fundamentals of Biostatistics. 8th ed. Boston, MA: Cengage Learning; 2015.
27. Vargas-Meza J, Cervantes-Armenta MA, Campos-Nonato I, Nieto C, Marrón-Ponce JA, Barquera S, et al. Dietary Sodium and Potassium Intake: Data from the Mexican National Health and Nutrition Survey 2016. Nutrients 2022 Jan 11;14(2):281. [doi: 10.3390/nu14020281]
28. Morales-Guerrero J, Miranda-Alatriste P, Villafuerte-Salazar M, Espinosa-Cuevas, Cassis-Nosthas L, Colín-Ramírez E. Determination of the chemical compositions of Mexican antojitos and dishes in Mexico City. Journal of Food Composition and Analysis 2023 May;118:105156 [FREE Full text] [doi: 10.1016/j.jfca.2023.105156]
29. Bhat S, Marklund M, Henry M, Appel L, Croft K, Neal B. A Systematic Review of the Sources of Dietary Salt Around the World. Adv Nutr 2020 May 1;11(3):677-686 [FREE Full text] [doi: 10.1093/advances/nmz134]
30. WHO. Guideline: Potassium intake for adults and children. Geneva: World Health Organization (WHO); 2012.

31. Mente A, Irvine EJ, Honey RJD, Logan AG. Urinary Potassium Is a Clinically Useful Test to Detect a Poor Quality Diet. The Journal of Nutrition 2009 Apr;139(4):743-749. [doi: 10.3945/jn.108.098319]
32. Parpia AS, Goldstein MB, Arcand J, Cho F, L'Abbé MR, Darling PB. Sodium-Reduced Meat and Poultry Products Contain a Significant Amount of Potassium from Food Additives. Journal of the Academy of Nutrition and Dietetics 2018 May;118(5):878-885. [doi: 10.1016/j.jand.2017.10.025]
33. Picard K, Picard C, Mager DR, Richard C. Potassium content of the American food supply and implications for the management of hyperkalemia in dialysis: An analysis of the Branded Product Database. Seminars in Dialysis 2024;37(4):307-316. [doi: 10.1111/sdi.13007]
34. WHO. The SHAKE Technical Package for Salt Reduction. The SHAKE Technical Package for Salt Reduction. Geneva, Switzerland: World Health Organization (WHO); 2016.
35. Hyseni L, Elliot-Green A, Lloyd-Williams F, Kypridemos C, O'Flaherty M, McGill R, et al. Systematic review of dietary salt reduction policies: Evidence for an effectiveness hierarchy? PLoS ONE 2017 May 18;12(5):e0177535. [doi: 10.1371/journal.pone.0177535]
36. Chavez Reynoso IJ, Ochoa Estrada R, Valenzuela Paredes L, Gayosso García Sancho LE, Tolano Villaverde IDJ. Actualización de la Norma Oficial Mexicana NOM-051-SCFI/SSA1 en la Industria de Alimentos y Bebidas. Jour. Bol. Cien 2021 Apr 23;17(Especial):6-18. [doi: 10.52428/20758944.v17iespecial.1]


## Abbreviations

**CVD:** cardiovascular disease
**DEI:** daily energy intake
**ISAK:** International Society for the Advancement of Kinanthropometry
**KAB:** knowledge, attitudes, and behaviors
**PAHO:** Pan American Health Organization
**WHO:** World Health Organization


*Edited by T de Azevedo Cardoso; submitted 10.02.24; peer-reviewed by X Liao, G Martin-Alemany; comments to author 14.08.24; revised version received 29.08.24; accepted 03.10.24; published 18.11.24.*

*Please cite as:*
*Gutiérrez-Salmeán G, Miranda-Alatriste PV, Benítez-Alday P, Orozco-Rivera LE, Islas-Vargas N, Espinosa-Cuevas Á, Correa-Rotter R, Colin-Ramirez E*
*Knowledge, Attitudes, and Behaviors Toward Salt Consumption and Its Association With 24-Hour Urinary Sodium and Potassium Excretion in Adults Living in Mexico City: Cross-Sectional Study*
*Interact J Med Res 2024;13:e57265*
*URL: https://www.i-jmr.org/2024/1/e57265*
*doi:10.2196/57265*
*PMID:39556832*

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 Deji[1,2*], MSc; Yuantao Tong[1*], PhD; Honglian Huang[1,3], PhD; Zeyu Zhang[1], PhD; Meng Fang[4], PhD; M James C Crabbe[5,6,7], PhD; Xiaoyan Zhang[1], PhD; Ying Wang[3], PhD

[1]Research Center for Translational Medicine, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China

[2]Information School, The University of Sheffield, Sheffield, United Kingdom

[3]Department of Clinical Laboratory Medicine Center, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China

[4]Department of Laboratory Medicine, Shanghai Eastern Hepatobiliary Surgery Hospital, Shanghai, China

[5]Wolfson College, Oxford University, Oxford, United Kingdom

[6]Institute of Biomedical and Environmental Science & Technology, University of Bedfordshire, Bedfordshire, United Kingdom

[7]School 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

***(Interact J Med Res 2024;13:e59070)*** 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.

*Submitted 01.04.24; this is a non–peer-reviewed article;accepted 03.04.24; published 10.04.24.*

*Please cite as:*
*Deji Z, Tong Y, Huang H, Zhang Z, Fang M, Crabbe MJC, Zhang X, Wang Y*
*Correction: 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:e59070*
*URL: https://www.i-jmr.org/2024/1/e59070*
*doi:10.2196/59070*
*PMID:38598823*

Corrigenda and Addenda

# Correction: Automated Psychotherapy in a Spaceflight Environment: Advantages, Drawbacks, and Unknowns

Logan Smith[1], PhD

Oklahoma State University, Stillwater, OK, United States

**Corresponding Author:**
Logan Smith, PhD
Oklahoma State University
306 Psychology Building
Stillwater, OK, 74078
United States
Phone: 1 772 242 5012
Email: logan.smith12@okstate.edu

**Related Article:**

Correction of: https://www.i-jmr.org/2024/1/e58803

***(Interact J Med Res 2024;13:e67671)*** doi:10.2196/67671

In "Automated Psychotherapy in a Spaceflight Environment: Advantages, Drawbacks, and Unknowns" (Interact J Med Res 2024;13:e58803) the author noted one error.

The following sentence was erroneously included in the *Acknowledgements* section:

> *LS is now affiliated with Austin Peay State University and is no longer at Oklahoma State University.*

This sentence has now been deleted, and the *Acknowledgments* section now reads as follows:

> *The author would like to thank Dr Nick Kanas for providing the inspiration for this review. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.*

The correction will appear in the online version of the paper on the JMIR Publications website on October 25, 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.

*Submitted 17.10.24; this is a non–peer-reviewed article;accepted 21.10.24; published 25.10.24.*

*Please cite as:*
*Smith L*
*Correction: Automated Psychotherapy in a Spaceflight Environment: Advantages, Drawbacks, and Unknowns*
*Interact J Med Res 2024;13:e67671*
*URL: https://www.i-jmr.org/2024/1/e67671*
*doi:10.2196/67671*
*PMID:*

Original Paper

# Establishment and Evaluation of a Noninvasive Metabolism-Related Fatty Liver Screening and Dynamic Monitoring Model: Cross-Sectional Study

Jiali Ni[1*], MD; Yong Huang[2*], MS; Qiangqiang Xiang[1*], MS; Qi Zheng[1], MS; Xiang Xu[2], MS; Zhiwen Qin[2], MS; Guoping Sheng[2], MD; Lanjuan Li[3], MD

[1]The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2]Shulan (Hangzhou) Hospital Affiliated to Zhejiang Shuren University, Shulan International Medical College, Hangzhou, China

[3]State Key Laboratory for Diagnosis and Treatment of Infectious Disease, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[*]these authors contributed equally

**Corresponding Author:**
Lanjuan Li, MD
State Key Laboratory for Diagnosis and Treatment of Infectious Disease
The First Affiliated Hospital
Zhejiang University School of Medicine
79 Qingchun Rd
Hangzhou, 310058
China
Phone: 86 571 87236458
Fax: 86 571 87236459
Email: ljli@zju.edu.cn

## Abstract

**Background:** Metabolically associated fatty liver disease (MAFLD) insidiously affects people's health, and many models have been proposed for the evaluation of liver fibrosis. However, there is still a lack of noninvasive and sensitive models to screen MAFLD in high-risk populations.

**Objective:** The purpose of this study was to explore a new method for early screening of the public and establish a home-based tool for regular self-assessment and monitoring of MAFLD.

**Methods:** In this cross-sectional study, there were 1758 eligible participants in the training set and 200 eligible participants in the testing set. Routine blood, blood biochemistry, and FibroScan tests were performed, and body composition was analyzed using a body composition instrument. Additionally, we recorded multiple factors including disease-related risk factors, the Forns index score, the hepatic steatosis index (HSI), the triglyceride glucose index, total body water (TBW), body fat mass (BFM), visceral fat area, waist-height ratio (WHtR), and basal metabolic rate. Binary logistic regression analysis was performed to explore the potential anthropometric indicators that have a predictive ability to screen for MAFLD. A new model, named the MAFLD Screening Index (MFSI), was established using binary logistic regression analysis, and BFM, WHtR, and TBW were included. A simple rating table, named the MAFLD Rating Table (MRT), was also established using these indicators.

**Results:** The performance of the HSI (area under the curve [AUC]=0.873, specificity=76.8%, sensitivity=81.4%), WHtR (AUC=0.866, specificity=79.8%, sensitivity=80.8%), and BFM (AUC=0.842, specificity=76.9%, sensitivity=76.2%) in discriminating between the MAFLD group and non-fatty liver group was evaluated ($P$<.001). The AUC of the combined model including WHtR, HSI, and BFM values was 0.900 (specificity=81.8%, sensitivity=85.6%; $P$<.001). The MFSI was established based on better performance at screening MAFLD patients in the training set (AUC=0.896, specificity=83.8%, sensitivity=82.1%) and was confirmed in the testing set (AUC=0.917, specificity=89.8%, sensitivity=84.4%; $P$<.001).

**Conclusions:** The novel MFSI model was built using WHtR, BFM, and TBW to screen for early MAFLD. These body parameters can be easily obtained using a body fat scale at home, and the mobile device software can record specific values and perform calculations. MFSI had better performance than other models for early MAFLD screening. The new model showed strong power and stability and shows promise in the area of MAFLD detection and self-assessment. The MRT was a practical tool to assess disease alterations in real time.

*(Interact J Med Res 2024;13:e56035)* doi:10.2196/56035

**KEYWORDS**

metabolic-associated fatty liver disease; nonalcoholic fatty liver disease; nonalcoholic steatohepatitis; body fat mass; waist-height ratio; basal metabolic rate; liver

## Introduction

Nonalcoholic fatty liver disease (NAFLD) is regarded as an important cause of liver disease, affecting more than 25% of the general population worldwide; more than 50% of patients with NAFLD also have dysmetabolism [1,2]. In 2020, experts redefined NAFLD as metabolically associated fatty liver disease (MAFLD), and much emphasis was placed on the presence of metabolic-related diseases or dysfunction [3-5]. Researchers have found that MAFLD is a multisystem disease, and liver steatosis is associated with type 2 diabetes, chronic kidney disease, cardiovascular disease, and other diseases that interact and form a vicious cycle [6-14]. China has the highest incidence of NAFLD or MAFLD morbidity in Asia [3,15,16]. Therefore, much attention should be given to MAFLD by enhancing awareness of MAFLD and optimizing its management.

To date, guidelines have suggested that liver biopsy could serve as the gold standard to diagnose histological liver damage, but noninvasive, quantitative assessment of liver fibrosis may also have prognostic implications. Ratziu et al [17] collected liver biopsy samples from 51 patients and found that 41% of the patients were at different stages of liver fibrosis or had nonalcoholic steatohepatitis. The uneven distribution of histological lesions inevitably led to sampling error when performing biopsy. Abdominal imaging, such as B-ultrasound imaging and the controlled attenuation parameter (CAP) technique, can be used to diagnosis liver disease; the former is less sensitive to mild steatosis, while the latter can detect steatosis of more than 5% and is one of the most common noninvasive methods for quantifying hepatic steatosis and fibrosis clinically [18,19]. The European Association for the Study of the Liver, European Association for the Study of Diabetes, and European Association for the Study of Obesity updated the clinical practice guidelines that propose that the nonalcoholic fatty liver disease fibrosis score (NFS) and fibrosis-4 (FIB-4) index can be used as prognostic markers for the progression of liver disease [20]. The NFS has higher specificity in the older adult population (individuals aged >65 years old) [21,22]. The predictive performance of the NFS, FIB-4 index, and aspartate aminotransferase-to-platelet ratio index (APRI) has been consistent in relation to rates of liver-related disease and mortality but is less valuable for the prediction of liver fibrosis [23]. One study found that the combination of the NFS, FIB-4 index, and liver stiffness measurement greatly improved the diagnostic accuracy, and the performance was similar to that of liver biopsy [24]. A cross-sectional study found that the triglyceride glucose (TyG) index was positively correlated with the likelihood and severity of NAFLD. The TyG index is generally considered a biomarker of steatosis, while its causal role in the judgement of fibrosis progression remains unclear [25,26]. In addition, the hepatic steatosis index (HSI) is more accurate in discriminating between MAFLD and nonfatty liver disease (non-FLD) than ultrasound. The predictive ability of the CAP for steatosis is superior to that of the HSI, and the HSI is more effective at discriminating patients with moderate-to-severe disease [18,27].

Studies have shown that numerous anthropometric indicators, such as BMI, waist-height ratio (WHtR), waist-hip ratio, and body adiposity index, are applicable for the quantification of visceral steatosis [28-32]. Body fat scales, a new popular domestic tool for health analysis, can be used to analyze basic parameters of body conditions such as the basal metabolic rate (BMR), body water distribution, and fat distribution. Reputable experts in the field have conducted extensive long-term studies on NAFLD and MAFLD, yet few noninvasive scoring models that accurately reflect disease activity or progression have been identified [33,34].

Therefore, there is an urgent need to identify more accurate predictive indicators and develop new screening methods for early MAFLD screening. The aim of this study was to construct a noninvasive prediction system for MAFLD, explore this new system for early screening in public, and establish a home-based tool for regular self-assessment and monitoring of MAFLD.

## Methods

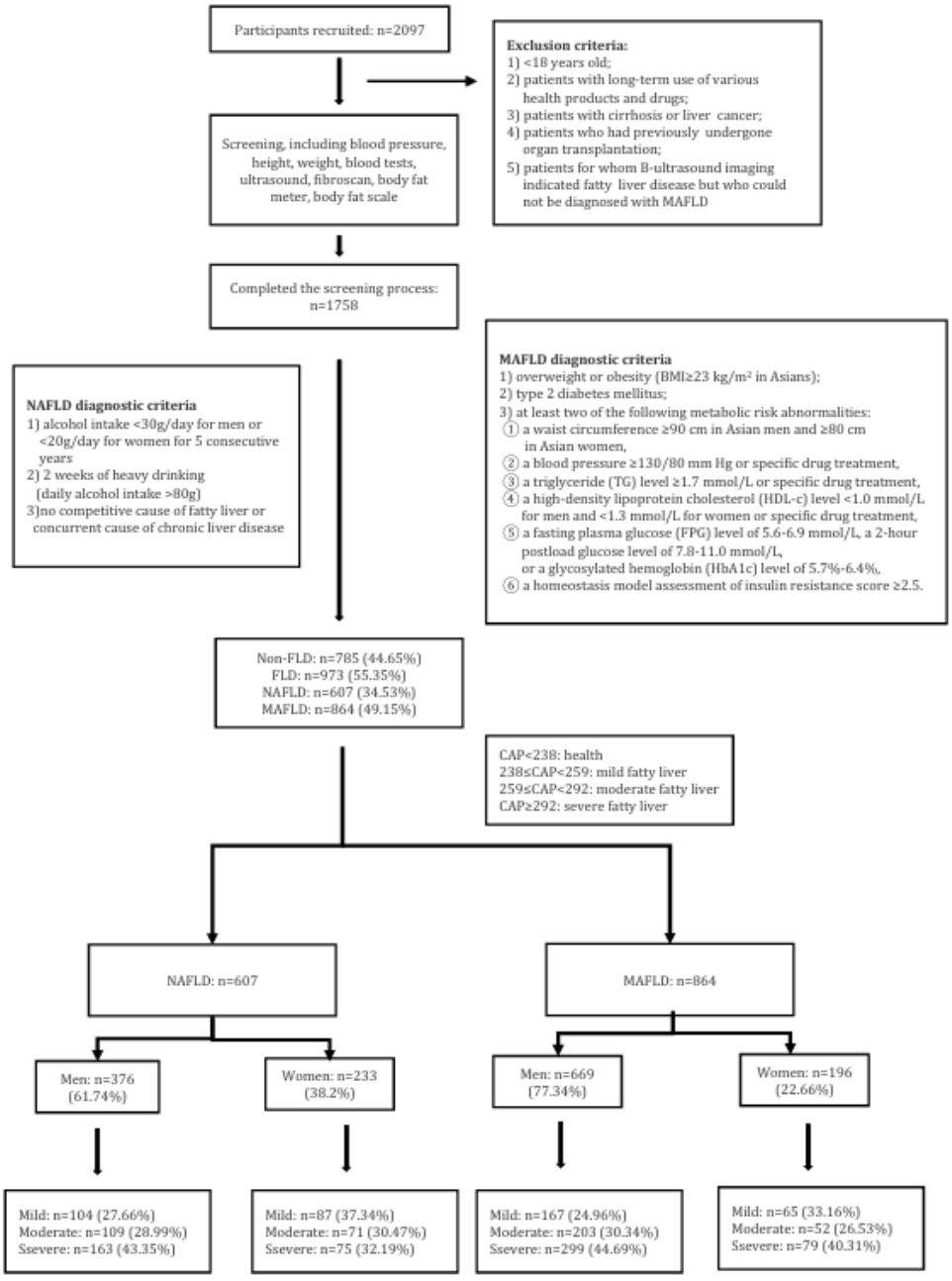
### Study Population

The participants came from Hangzhou, Shaoxing, and Quzhou from March 2021 to November 2021, and a total of 2097 participants were enrolled (Figure 1). All participants signed the informed consent form and completed the examination as required. There were 1758 eligible participants in the training set who truthfully and completely answered the questionnaire, which contained items regarding height, weight, drinking history, past medical history, and other basic information.

To validate the results of the training set, there were 200 eligible participants grouped into the testing set.

All participants were diagnosed using the liver stiffness measurement and classified according to the CAP. CA*P* values <238 was considered to indicate a healthy liver, ≥238 and <259 was considered to indicate mildly fatty liver, ≥259 and <292 was considered to indicate moderately fatty liver, and ≥292 was considered to indicate severely fatty liver [35,36].

**Figure 1.** Flowchart of the inclusion process for the participants in the training set and the proportion of mild, moderate, and severe fatty liver disease in the nonalcoholic fatty liver disease (NAFLD) group and metabolically associated fatty liver disease (MAFLD) group. CAP: controlled attenuation parameter; FLD: fatty liver disease.




## Exclusion Criteria

Patients who met one or more of the following criteria could not participate in this study: (1) <18 years old; (2) long-term use of various health products and drugs; (3) presence of cirrhosis or liver cancer; (4) previous organ transplantation; and (5) patients for whom B-ultrasound imaging indicated FLD but who could not be diagnosed with MAFLD.

## Diagnostic Criteria

The researchers in this study entered and organized the data, and the following diagnostic criteria were used to distinguish MAFLD patients [3]: (1) overweight or obese (BMI ≥23 kg/m$^2$ in Asians), (2) presence of type 2 diabetes mellitus, and (3) at least 2 of the following metabolic risk abnormalities: waist circumference ≥90 cm in Asian men and ≥80 cm in Asian women; blood pressure ≥130/80 mm Hg or specific drug treatment; triglyceride (TG) level ≥1.7 mmol/L or specific drug treatment; high-density lipoprotein cholesterol (HDL-c) level <1.0 mmol/L for men and <1.3 mmol/L for women or specific drug treatment; fasting plasma glucose (FPG) level of 5.6 mmol/L to 6.9 mmol/L, 2-hour postload glucose level of 7.8 mmol/L to 11.0 mmol/L, or glycosylated hemoglobin level of 5.7% to 6.4%; and homeostasis model assessment of insulin resistance score ≥2.5.

## Data Collection and Model Selection

All items were completed under the guidance of the researchers. The participants underwent fasting blood tests. Waist circumference and hip circumference were measured with the

participants wearing thin clothes. Body composition analysis was performed with bare feet. The patients were in a supine position during the FibroScan exam, and the right upper limb was held high and flat close to the ear. The probe was moved a small distance from the anchor point so that the most suitable detection point could be determined.

We collected basic information, including sex, age, height, weight, BMI, waist circumference, hip circumference, blood pressure, heart rate, and alcohol consumption history. The following laboratory results were included: alanine aminotransferase (ALT), aspartate aminotransferase (AST), glutamyl transpeptidase (GGT), alkaline phosphatase, hemoglobin, total cholesterol (TC), TG, HDL-c, low-density lipoprotein cholesterol (LDL-c), uric acid, and FPG levels, as well as white blood cell, red blood cell, and platelet (PLT) counts.

A body composition analyzer (InBody770, Biospace) was used to measure body composition and determine total body water (TBW), intracellular water, skeletal muscle mass, protein, and body fat mass (BFM). A body fat scale (3 Pro, Huawei) was used to determine the BMR, fat%, and visceral fat area (VFA).

The models or formulas involved in this study, including BMI, FIB-4 index, Forns index score, APRI, glutamyl transpeptidase-to-platelet ratio index (GPR), HSI, and TyG index, were developed using the following standard equations:

$$BMI=weight/height^2$$

$$FIB\text{-}4=age\times AST/(PLT\times\sqrt{ALT})$$

$$\text{Forns index score}=7.811-3.131\times \ln PLT(109/L)+0.781\times \ln GGT+3.467\times \ln age-0.014\times TC$$

$$APRI=(AST/\text{upper limit of normal})/PLT\times 100$$

$$\text{glutamyl transpeptidase-to-platelet ratio index}=(GGT/\text{upper limit of normal})/PLT\times 100$$

$$HSI=8\times(ALT/AST)+BMI\ (\text{female}+2,\ \text{diabetes}+2)$$

$$TyG=\ln(TG\times FPG/2)$$

### Statistical Analysis Methods

Participants were divided into the non-FLD group, which was the healthy group; MAFLD group; and NAFLD group.

All data obtained in this study were analyzed using SPSS version 26.0 (IBM Corp). The continuous variables were tested for normality and homogeneity of variance. A *t* test was performed for measurement data that followed a normal distribution, and the results are expressed as the mean (SD). Nonnormally distributed data were analyzed using nonparametric tests, and the results are represented by quartiles. The chi-square test or Fisher precision probability test was used for quantitative data such as sex. ANOVA was followed by post hoc analysis tests to compare numerical data among the 3 groups (MAFLD, NAFLD, and non-FLD). A $P<.01$ indicated that the difference was statistically significant.

Binary logistic regression analysis was performed to explore the potential anthropometric indicators with predictive ability to screen for MAFLD. A receiver operating characteristic (ROC) curve was drawn based on the selected indicators, and the area under the ROC curve (AUC) was calculated correspondingly. The indicator with the highest AUC was considered the most valuable indicator. The maximum Youden index (using the formula sensitivity + specificity - 1) was used to define the optimal cutoff value. Potential confounding variables were added into the logistic regression equation step by step, including age; blood pressure; and FPG, TC, TG, HDL-c, and LDL-c levels. Calibration Model I (age, blood pressure, and FPG level were added to the logistic regression equation) and Model II (age; blood pressure; and FPG, TC, TG, HDL-c, and LDL-c levels were added to the logistic regression equation) were established, and the predictive ability was evaluated before and after calibration. All significant indicators were included for the combination of diagnostic tests, and ROC curves were drawn. A new prediction model, the MAFLD Screening Index (MFSI), was constructed using logistic regression analysis, and the model was validated with the testing set. All tests were 2-tailed, and $P<.01$ was considered statistically significant.

### Ethical Considerations

Every participant signed a written informed consent form and participated in the study anonymously. We ensured it was not possible to identify individual participants in any images used in manuscripts or other materials.

Every participant was given an allowance of ¥300 (US $0.14) upon completion of the research project. The study protocol was approved by the Ethics Committee of Shulan Hangzhou Hospital (approval number KY2021001).

The study did not involve additional invasive procedures, and there were no associated adverse reactions.

## Results

### Comparing Numerical Data Among the 3 Groups (MAFLD, NAFLD, and Non-FLD)

Using ANOVA to compare the basic information and anthropometric indicators among the 3 groups, the results showed that all parameters were significantly different among the MAFLD, NAFLD, and non-FLD groups. After the post hoc analysis, PLT count ($P=.10$) was not significantly different between the MAFLD and non-FLD groups, and white blood cell count ($P=.26$), TC ($P=.35$), LDL-c ($P=.11$), VFA ($P=.07$), and Fat% ($P=.38$) were not significantly different between the MAFLD and NAFLD groups (Table 1).

**Table 1.** Baseline characteristics and anthropometric indicators compared among 3 groups: metabolically associated fatty liver disease (MAFLD), nonalcoholic fatty liver disease (NAFLD), and non-fatty liver disease (non-FLD)

| Characteristics | non-FLD (n=786) | MAFLD (n=864) | NAFLD (n=607) | Statistic (*df*) | *P* value |
|---|---|---|---|---|---|
| **Sex, n (%)** | | | | 224.985[a] (2) | <.001 |
| Male | 326 (41.5) | 668 (77.3) | 375 (61.8) | | |
| Female | 460 (58.5) | 196 (22.7) | 232 (38.2) | | |
| Age (years), mean (range) | 36 (28-48) | 45 (34-55) | 40 (31-53) | 107.212[b] (2) | <.001 |
| Height (cm), mean (SD) | 164.17 (7.820) | 168.23 (7.844) | 166.81 (8.358) | 54.873[c] | <.001 |
| Weight (kg), mean (range) | 57.30 (51.60-64.30) | 72.9 (66.2-80.7) | 69.7 (61.1-78.1) | 664.463[b] (2) | <.001 |
| BMI (kg/m$^2$), mean (range) | 21.49 (20.06-23.18) | 25.66 (24.04-27.79) | 25.04 (22.99-2t7.12) | 798.113[b] (2) | <.001 |
| SBP[d] (mm Hg), mean (range) | 119 (110-135) | 132 (121-144) | 128 (118-140) | 218.758[b] (2) | <.001 |
| DBP[e] (mm Hg), mean (SD) | 74.89 (11.216) | 82.88 (11.863) | 79.72 (11.611) | 91.652[c] (2,2249) | <.001 |
| WBC[f] (109/L), mean (range) | 5.80 (4.95-6.80) | 6.5 (5.5-7.6) | 6.3 (5.4-7.5) | 91.944[b] (2) | <.001 |
| RBC[g] (1012/L), mean (range) | 4.69 (4.39-5.11) | 5.12 (4.81-5.42) | 5.02 (4.65-5.38) | 210.018[b] (2) | <.001 |
| Hb[h] (g/L), mean (range) | 140 (131-153) | 155 (145-163) | 150 (137-160) | 213.919[b] (2) | <.001 |
| PLT[i] (109/L), mean (SD) | 234.66 (55.331) | 238.18 (58.518) | 244.67 (58.671) | 5.041[c] (2,2091) | .006 |
| FPG[j] (mmol/L), mean (range) | 4.60 (4.36-4.89) | 4.88 (4.55-5.34) | 4.80 (4.50-5.19) | 128.314[b] (2) | <.001 |
| ALT[k] (U/L), mean (range) | 14 (10-20) | 27.00 (19.00-34.00) | 24.00 (17.75-40.00) | 475.850[b] (2) | <.001 |
| AST[l] (U/L), mean (range) | 20 (17-24) | 25.00 (21.00-32.00) | 24.00 (19.00-29.00) | 245.274[b] (2) | <.001 |
| ALP[m] (U/L), mean (range) | 57 (47-70) | 69.00 (57.00-83.00) | 66.00 (55.00-81.00) | 146.960[b] (2) | <.001 |
| GGT[n] (U/L), mean (range) | 15 (12-21) | 31.00 (20.00-52.00) | 24.00 (17.00-38.00) | 496.679[b] (2) | <.001 |
| TC[o] (mmol/L), mean (range) | 4.71 (4.13-5.23) | 4.99 (4.38-5.33) | 4.96 (4.33-5.55) | 47.740[b] (2) | <.001 |
| TG[p] (mmol/L), mean (range) | 0.93 (0.71-1.26) | 1.76 (1.21-2.45) | 1.49 (1.07-2.24) | 529.457[b] (2) | <.001 |
| HDL-c[q] (mmol/L), mean (range) | 1.48 (1.28-1.68) | 1.21 (1.07-1.40) | 1.26 (1.07-1.45) | 284.363[b] (2) | <.001 |
| LDL-c[r] (mmol/L), mean (range) | 2.64 (2.23-3.16) | 3.10 (2.64-3.64) | 3.04 (2.56-3.55) | 146.926[b] (2) | <.001 |
| UA[s] (μmol/L), mean (SD) | 295.99 (80.86) | 378.82 (85.081) | 358.39 (87.676) | 206.430[c] (2,2234) | <.001 |
| E-value (kPa), mean (range) | 4.40 (3.70-5.20) | 5.2 (4.3-6.3) | 5.0 (4.2-6.0) | 164.275[b] (2) | <.001 |
| CAP[t] (dB/m), mean (range) | 200 (179-218) | 284 (257-320) | 278 (255-315) | 1538.285[b] (2) | <.001 |
| WHtR[u], mean (SD) | 0.459 (0.0656) | 0.524 (0.5117) | 0.512 (0.0984) | 113.769[c] | <.001 |
| WHR[v], mean (range) | 0.829 (0.783-0.880) | 0.918 (0.872-0.950) | 0.892 (0.840-0.939) | 284.420[b] (2) | <.001 |
| FIB-4[w] index, mean (range) | 0.813 (0.588-1.234) | 0.954 (0.608-1.362) | 0.782 (0.521-1.235) | 15.698[b] (2) | <.001 |
| Forns index, mean (SD) | 5.471 (1.6797) | 6.536 (1.5977) | 6.005 (1.6630) | 166.437[c] (2,2190) | <.001 |
| APRI[x], mean (range) | 0.240 (0.187-0.301) | 0.282 (0.215-0.376) | 0.263 (0.201-0.351) | 72.070[b] (2) | <.001 |
| GPR[y], mean (range) | 0.133 (0.103-0.192) | 0.230 (0.159-0.403) | 0.186 (0.135-0.292) | 325.345[b] (2) | <.001 |
| HSI[z], mean (rang) | 28.57 (26.31-30.84) | 35.28 (32.02-39.04) | 34.47 (30.87-38.63) | 772.797[b] (2) | <.001 |
| TyG[aa] index, mean (range) | 8.137 (7.863-8.463) | 8.855 (8.463-9.234) | 8.674 (8.322-9.092) | 585.116[b] (2) | <.001 |
| TBW[bb] (kg), mean (range) | 30.30 (27.2-036.40) | 38.65 (33.80-42.90) | 36.85 (30.23-42.88) | 341.198[b] (2) | <.001 |
| ICW[cc], mean (range) | 18.80 (16.70-22.60) | 24.10 (20.90-26.80) | 22.80 (18.70-26.10) | 340.014[b] (2) | <.001 |

| Characteristics | non-FLD (n=786) | MAFLD (n=864) | NAFLD (n=607) | Statistic (*df*) | *P* value |
|---|---|---|---|---|---|
| Protein, mean (range) | 8.10 (7.20-9.80) | 10.40 (9.08-11.60) | 9.90 (8.10-11.30) | 339.536[b] (2) | <.001 |
| BFM[dd], mean (range) | 14.90 (12.00-17.40) | 20.80 (17.60-24.43) | 20.10 (16.80-24.20) | 638.768[b] (2) | <.001 |
| SMM[ee], mean (range) | 22.50 (19.80-27.53) | 29.40 (25.30-32.93) | 27.80 (22.35-32.10) | 336.593[b] (2) | <.001 |
| BMR[ff] (kcal), mean (range) | 1283.95 (1179.70-1450.17) | 1526.55 (1387.09-1655.52) | 1470.15 (1276.65-1622.62) | 358.763[b] (2) | <.001 |
| VFA[gg], mean (range) | 6.552 (5.235-7.762) | 9.013 (7.598-10.813) | 8.753 (7.294-10.671) | 518.559[b] (2) | <.001 |
| Fat (%), mean (SD) | 25.319 (5.9566) | 28.255 (5.2256) | 28.524 (5.7358) | 68.099[c] (2,2018) | <.001 |

[a]Chi-squared test.
[b]*H* value.
[c]*F* value.
[d]SBP: systolic blood pressure.
[e]DBP: diastolic blood pressure.
[f]WBC: white blood cell.
[g]RBC: red blood cell.
[h]Hb: hemoglobin.
[i]PLT: platelet.
[j]FPG: fasting plasma glucose.
[k]ALT: alanine aminotransferase.
[l]AST: aspartate aminotransferase.
[m]ALP: alkaline phosphatase.
[n]GGT: glutamyl transpeptidase.
[o]TC: total cholesterol.
[p]TG: triglyceride.
[q]HDL-c: high-density lipoprotein cholesterol.
[r]LDL-c: low-density lipoprotein cholesterol.
[s]UA: uric acid.
[t]CAP: controlled attenuation parameter.
[u]WHtR: waist-height ratio.
[v]WHR: waist-hip ratio.
[w]FIB-4: fibrosis-4.
[x]APRI: aspartate aminotransferase-to-platelet ratio index.
[y]GPR: glutamyl transpeptidase-to-platelet ratio index.
[z]HSI: hepatic steatosis index.
[aa]TyG: triglyceride glucose.
[bb]TBW: total body water.
[cc]ICW: intracellular water.
[dd]BFM: body fat mass.
[ee]SMM: skeletal muscle mass.
[ff]BMR: basal metabolic rate.
[gg]VFA: visceral fat area.

## Predictive Performance of Different Anthropometric Indicators

The variables in the previous section with a $P<.01$ were further included in the logistic regression analysis. The ROC curves and optimal cutoff points for the selected indicators are shown in Table 2 and Figure 2. The AUCs of the WHtR, the Forns index, the HSI, the TyG index, TBW, BFM, and BMR were 0.866, 0.684, 0.873, 0.835, 0.760, 0.842, and 0.778, respectively ($P<.001$; Table 2).

**Table 2.** Cutoff points and areas under the curve (AUCs) were used to demonstrate the screening ability of the different anthropometric indicators for metabolically associated fatty liver disease (n=1649).

| Anthropometric indicators | AUC (95% CI) | *P* value | Cutoff point | Specificity (%) | Sensitivity (%) |
|---|---|---|---|---|---|
| WHtR[a] | 0.866 | <.001 | 0.501449713 | 79.8 | 80.8 |
| Forns index | 0.684 | <.001 | 6.160599276 | 67.8 | 61.3 |
| HSI[b] | 0.873 | <.001 | 31.15061285 | 76.8 | 81.4 |
| TyG[c] index | 0.835 | <.001 | 8.450341708 | 74 | 76.5 |
| TBW[d] (kg) | 0.760 | <.001 | 36.55 | 75.3 | 65.3 |
| BFM[e] | 0.842 | <.001 | 17.55 | 76.9 | 76.2 |
| BMR[f] | 0.778 | <.001 | 1434.263124 | 73.1 | 70.1 |
| Combination (WHtR/HSI) | 0.885 | <.001 | N/A[g] | 76 | 85.6 |
| Combination (WHtR/BFM) | 0.881 | <.001 | N/A | 81.7 | 80 |
| Combination (BFM/HSI) | 0.889 | <.001 | N/A | 81 | 82.9 |
| Combination (WHtR/HSI/BFM) | 0.900 | <.001 | N/A | 81.8 | 85.6 |

[a]WHtR: waist-to-hip ratio.

[b]HSI: hepatic steatosis index.

[c]TyG: triglyceride glucose.

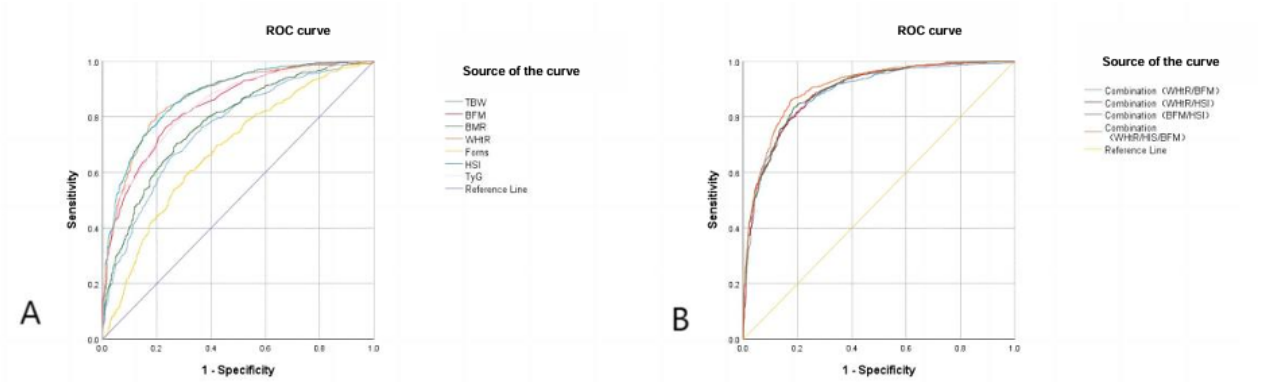
[d]TBW: total body water.

[e]BFM: body fat mass.

[f]BMR: basal metabolic rate.

[g]N/A: not applicable.

**Figure 2.** Receiver operating characteristic (ROC) curves for the screening ability of different anthropometric indicators for metabolically associated fatty liver disease (MAFLD): (A) screening ability of the waist-height ratio (WHtR), Frons index, hepatic steatosis index (HSI), triglyceride glucose (TyG) index, total body water (TBW), body fat mass (BFM), basal metabolic rate (BMR) and (B) screening ability of combinations of WHtR, HSI, and BFM. Diagonal segments were produced by ties.




According to the ROC curve and AUC (Figure 2), HSI had the strongest predictive performance for MAFLD in the training set, and the performance ranking was as follows: HSI, WHtR, BFM, TyG index, TBW, and Forns index. The confounding factors were further corrected for in the logistic regression analysis (Model I: age, blood pressure, and FPG level were added to the logistic regression equation; Model II: age; blood pressure; and FPG, TC, TG, HDL-c, and LDL-c levels were added to the logistic regression equation). After correction for confounding factors, the odds ratio (OR) of the Forns index in Model I was 1.043 (95% CI 0.851-1.277), and that in Model II was 1.050 (95% CI 0.854-1.293; Table 3). The results showed that the performance of the Forns index for MAFLD screening was unstable and the performance of other anthropometric indicators was not easily influenced by confounders.

The HSI and WHtR showed better predictive performance than the other indicators. The sensitivity of the HSI was higher than that of the other anthropometric indicators (sensitivity=81.4%), and the specificity of the WHtR was higher than that of the other anthropometric indicators (specificity=79.8%). The combination of WHtR, HSI, and BFM increased the predictive ability for MAFLD, and the AUC was 0.900 (specificity=81.8%, sensitivity=85.6%; *P*<.001; Table 2).

**Table 3.** Confounders were corrected in the binary logistic regression analysis to compare changes in screening power for different anthropometric indicators (n=1649).

| Anthropometric indicators | Nonadjusted model | | Model I | | Model II | |
|---|---|---|---|---|---|---|
| | OR[a] (95% CI) | *P* value | OR (95% CI) | *P* value | OR (95% CI) | *P* value |
| TBW[b] (kg) | 1.079 (1.051-1.109) | <.001 | 1.086 (1.055-1.118) | <.001 | 1.075 (1.042-1.108) | <.001 |
| BFM[c] | 1.255 (1.194-1.319) | <.001 | 1.250 (1.186-1.317) | <.001 | 1.257 (1.192-1.326) | <.001 |
| WHtR[d] | 145.540 (9.015-2349.524) | <.001 | 107.825 (6.232-1865.456) | .001 | 113.408 (6.759-1902.900) | <.001 |
| Forns index | 1.166 (1.053-1.292) | .003 | 1.043 (0.851-1.277) | .69 | 1.050 (0.854-1.293) | .64 |
| HSI[e] | 1.124 (1.070-1.182) | <.001 | 1.128 (1.070-1.190) | <.001 | 1.110 (1.052-1.172) | <.001 |
| TyG[f] index | 6.557 (4.560-9.427) | <.001 | 5.832 (3.972-8.562) | <.001 | 6.005 (3.764-9.579) | <.001 |

[a]OR: odds ratio.

[b]TBW: total body water.

[c]BFM: body fat mass.

[d]WHtR: waist-height ratio.

[e]HSI: hepatic steatosis index.

[f]TyG: triglyceride glucose.

## Development of a New MAFLD Screening Model

The HSI, WHtR, and BFM displayed strong power in screening for MAFLD. The HSI was calculated based on BMI, ALT levels, and AST levels and was not suitable for early screening for MAFLD. The purpose of establishing a new model was to reduce the need for invasive procedures and reduce the frequency of medical visits, as well as to screen for MAFLD in high-risk populations. The predictive ability of TBW was stable after correcting for confounders (Model I: 95% CI 1.055-1.118; Model II: 95% CI 1.042-1.108; Table 3). Therefore, TBW was included in the new model. Logistic regression analysis was used to establish the MAFLD early screening model, which was named the MFSI. The formula was as follows: MFSI=–13.968+0.120×TBW+0.254×BFM+10.793×WHtR (Figure 3). The AUC of the MFSI was 0.896 (specificity: 83.8%, sensitivity: 82.1%; *P*<.001; Table 4). Collectively, the performance of the MFSI and the WHtR/HSI/BFM combination models was similar.

**Figure 3.** Receiver operator characteristic (ROC) curves showing the screening ability of different combinations of anthropometric indicators and a new metabolically associated fatty liver disease (MAFLD) screening model named the MAFLD screening index (MFSI=–13.968+0.120×total body water [TBW]+0.254×body fat mass [BFM]+10.793×waist-height ratio [WHtR]). Diagonal segments were produced by ties. BMR: basal metabolic rate; HSI: hepatic steatosis index; TyG: triglyceride glucose.

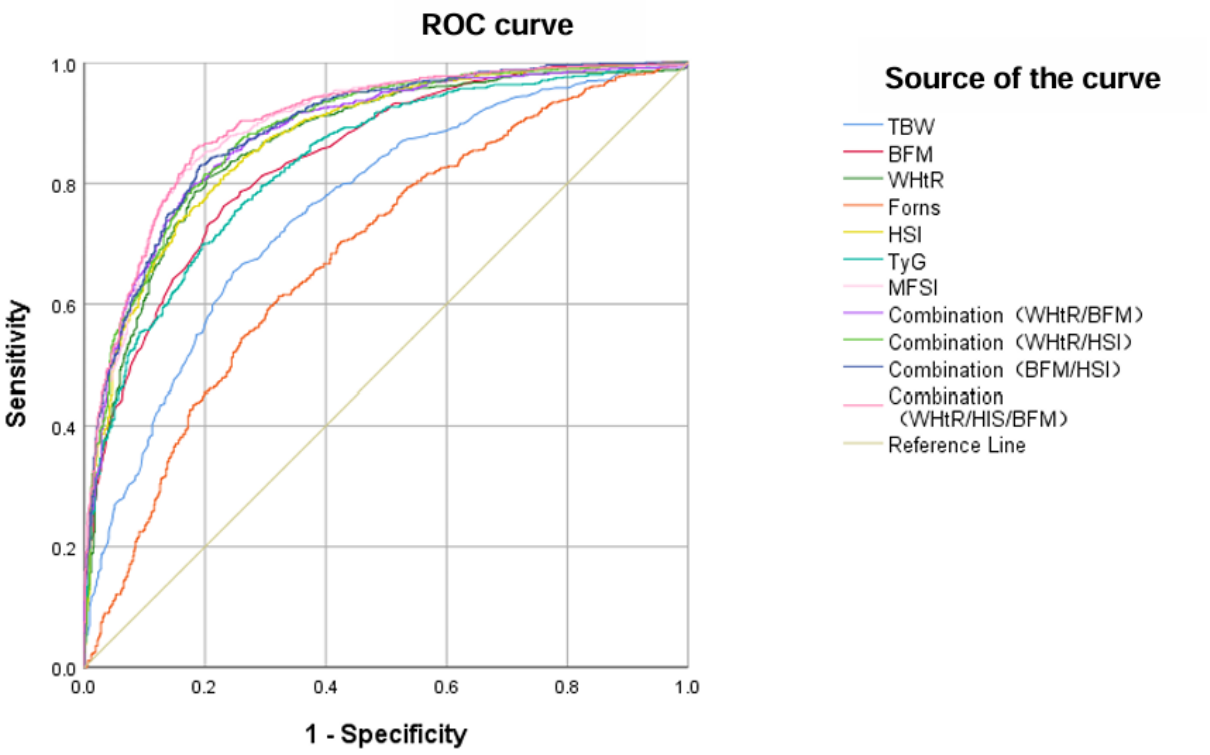



**Table 4.** Cutoff points and areas under the curve (AUCs) were used to compare the screening ability of different anthropometric indicators and the metabolically associated fatty liver disease screening index (MFSI; n=1649).

| Anthropometric indicators | AUC (95% CI) | *P* value | Cutoff point | Specificity (%) | Sensitivity (%) |
|---|---|---|---|---|---|
| WHtR[a] | 0.866 | <.001 | 0.501449713 | 79.8 | 80.8 |
| Forns index | 0.684 | <.001 | 6.160599276 | 67.8 | 61.3 |
| HSI[b] | 0.873 | <.001 | 31.15061285 | 76.8 | 81.4 |
| TyG[c] index | 0.835 | <.001 | 8.450341708 | 74 | 76.5 |
| TBW[d] (kg) | 0.760 | <.001 | 36.55 | 75.3 | 65.3 |
| BFM[e] | 0.842 | <.001 | 17.55 | 76.9 | 76.2 |
| Combination (WHtR/HSI) | 0.885 | <.001 | N/A[f] | 76 | 85.6 |
| Combination (WHtR/BFM) | 0.881 | <.001 | N/A | 81.7 | 80 |
| Combination (BFM/HSI) | 0.889 | <.001 | N/A | 81 | 82.9 |
| Combination (WHtR/HSI/BFM) | 0.900 | <.001 | N/A | 81.8 | 85.6 |
| MFSI | 0.896 | <.001 | 0.5146795 | 83.8 | 82.1 |

[a]WHtR: waist-height ratio.

[b]HSI: hepatic steatosis index.

[c]TyG: triglyceride glucose.

[d]TBW: total body water.

[e]BFM: body fat mass.

[f]N/A: not applicable.

### Performance of the MFSI in the Testing Set

There were a further 200 participants enrolled in the testing set, including 51 non-FLD patients and 149 MAFLD patients. To evaluate the predictive ability of the MFSI for screening for MAFLD in a high-risk population, the MFSI was used with the testing set, and ROC curves were drawn based on the MFSI; BFM; WHtR; HSI; TBW; and the combined model with WHtR, HSI, and BFM (Figure 4). The AUC (testing set) of the MFSI was 0.917, the specificity was 89.8%, and the sensitivity was 84.4%. The AUC (testing set) of the combined model with WHtR, HSI, and BFM was 0.920, the specificity was 89.8%, and the sensitivity was 81.6% (Table 5). The performance of the MFSI was similar to that of the combined model with WHtR, HSI, and BFM in the testing set.

**Figure 4.** Receiver operating characteristic (ROC) curves for the screening ability of different anthropometric indicators and metabolically associated fatty liver disease screening index (MFSI) in the testing set. Diagonal segments were produced by ties. BFM: body fat mass; HSI: hepatic steatosis index; TBW: total body water; WHtR: waist-height ratio.

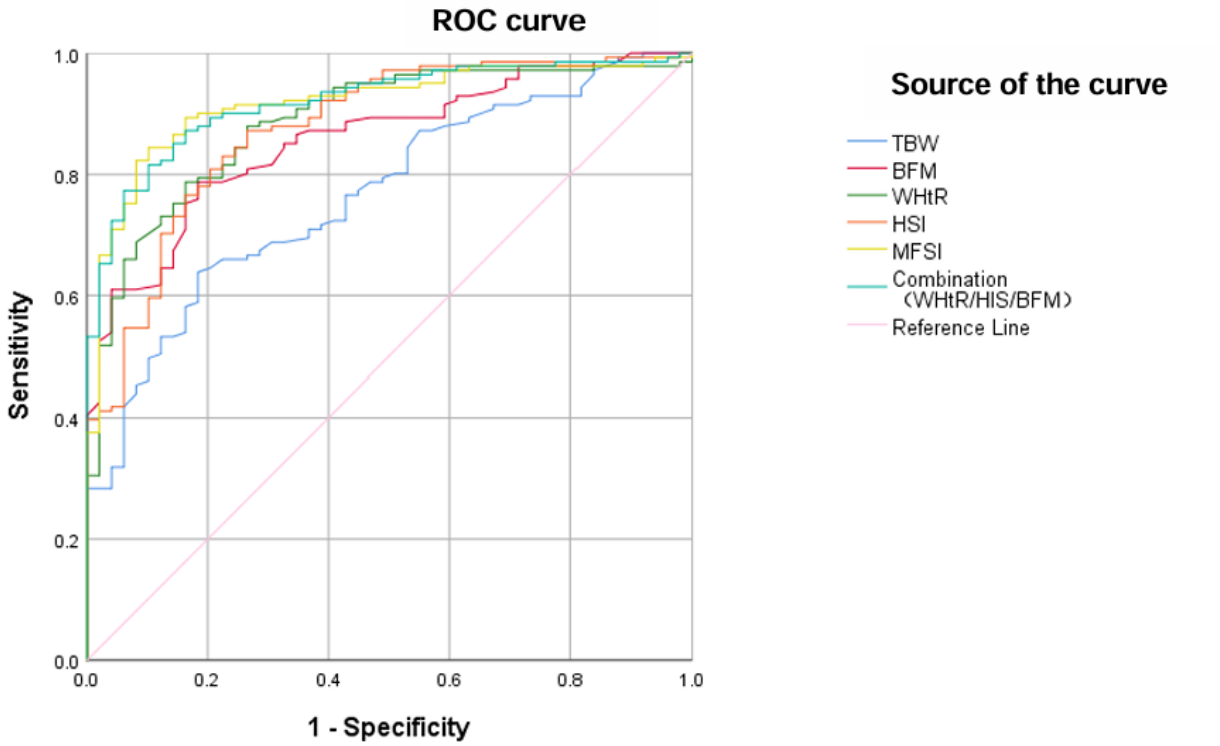



**Table 5.** Areas under the curve (AUCs) were used to compare the ability of the different anthropometric indicators and the metabolically associated fatty liver disease (MAFLD) screening index (MFSI) to screen for MAFLD in the testing set (n=200).

| Anthropometric indicators | AUC (95% CI) | *P* value | Specificity (%) | Sensitivity (%) |
|---|---|---|---|---|
| WHtR[a] | 0.886 | <.001 | 83.7 | 78.7 |
| TBW[b] (kg) | 0.767 | <.001 | 81.6 | 63.8 |
| BFM[c] | 0.858 | <.001 | 81.6 | 78.7 |
| HSI[d] | 0.877 | <.001 | 73.5 | 87.2 |
| Combination (WHtR/HSI/BFM) | 0.920 | <.001 | 89.8 | 81.6 |
| MFSI | 0.917 | <.001 | 89.8 | 84.4 |

[a]WHtR: waist-height ratio.

[b]TBW: total body water.

[c]BFM: body fat mass.

[d]HSI: hepatic steatosis index.

### MAFLD Rating Table for Prediction of MAFLD

The scoring system based on the MFSI and the application program was more practical for patient self-assessment. The MAFLD Rating Table (MRT) also included TBW, BFM, and WHtR. An MRT score ranging from 0 to 2 indicated a healthy individual, and a score ≥3 indicated MAFLD (Table 6). The

AUC of the MRT for MAFLD prediction was 0.876 ($P$<.001; Figure 5).

**Table 6.** Simple rating table to assess risk factors for metabolically associated fatty liver disease (MAFLD).

| Factors | Rating | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| TBW[a] (kg) | <33.35 | 33.35-45.05 | ≥45.05 | N/A[b] | N/A |
| BFM[c] | <17.55 | 17.55-20.15 | 10.15-22.95 | ≥22.95 | N/A |
| WHtR[d] | <0.501 | N/A | 0.501-0.525 | 0.525-0.538 | ≥0.538 |

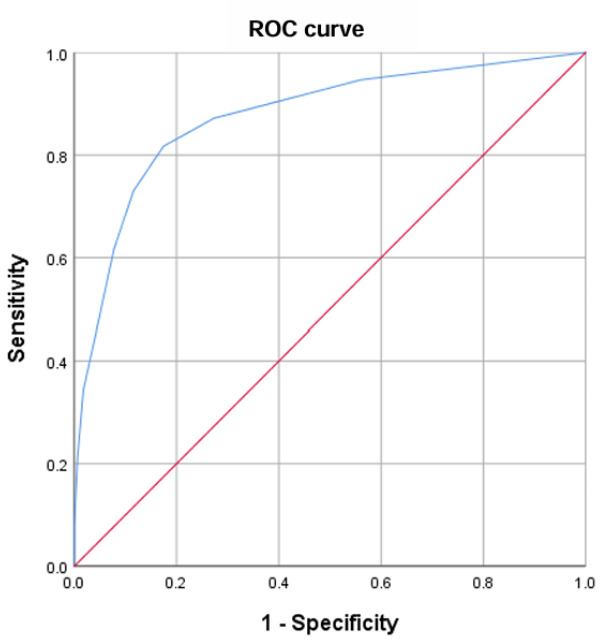
[a]TBW: total body water.

[b]N/A: not applicable.

[c]BFM: body fat mass.

[d]WHtR: waist-to-height ratio.

**Figure 5.** Receiver operating characteristic (ROC) curves of the correlation between the metabolically associated fatty liver disease (MAFLD) Rating Table (MRT) and MAFLD. Diagonal segments were produced by ties.




## Discussion

WHtR, BFM, and TBW were predictors of MAFLD. The AUC of the WHtR was 0.866 (specificity=79.8%, sensitivity=80.8%), the AUC of BFM was 0.842 (specificity=76.9%, sensitivity=76.2%), and the AUC of TBW was 0.760 (specificity=75.3%, sensitivity=65.3%). The novel MFSI model, derived through logistic regression analysis, included the WHtR, BFM, and TBW. Notably, the MFSI demonstrated independence from laboratory findings. Upon validation, the MFSI exhibited stability while offering advantages in terms of sensitivity and specificity for MAFLD screening (training set: AUC=0.896, specificity=83.8%, sensitivity=82.1%; testing set: AUC=0.917, specificity=89.8%, sensitivity=84.4%).

Researchers found that the measurement of visceral fat can predict the occurrence of chronic diseases, such as diabetes, hyperuricemia, and metabolic syndrome [37,38]. NAFLD and MAFLD affect more than 25% of the global population and are considered different stages of the disease course. Because of long-term subtle inflammation and unobvious clinical manifestations, some patients gradually develop liver fibrosis and cirrhosis [1]. It is important to raise awareness within the population and optimize the management of this disease.

In recent decades, researchers have considered that the APRI, FIB-4 index, BMI, HSI, and TyG index have high accuracy for the diagnosis of liver fibrosis. Nonfibrosis scores were higher in patients with MAFLD than in those with NAFLD [11,39-42]. Similar conclusions were drawn in this work. To distinguish patients with MAFLD in our study population, we compared traditional indicators and body composition between the MAFLD and non-FLD groups and found there was a significant difference between the 2 groups. Although traditional indicators had efficient performance for the prediction of liver fibrosis, it was doubtful that these indicators were robust for the screening of MAFLD before being confirmed by histological liver examination. Previously published work mainly focused on the predictive ability of indicators for the detection of liver fibrosis,

while much work omitted the performance of these indicators for the early screening of MAFLD.

Lee et al [27] suggested a new indicator named the HSI, and they found that NAFLD cannot be diagnosed when the HSI was <30.0, with a sensitivity of 92.5% (95% CI 91.4-93.5) The HSI showed similar performance for MAFLD diagnosis in this study. Italian researchers proposed another new indicator, the fatty liver index (FLI), which is calculated based on waist circumference, BMI, TG levels, and GGT levels. When the FLI is <30, a diagnosis of FLD can be ruled out, and when the FLI is ≥60, patients can be diagnosed with FLD. Waist circumference and BMI are the most robust predictors for the screening of FLD [43]. In contrast to the HSI, the FLI was established by incorporating waist circumference. However, BMI and waist circumference are totally different for people with varied dietary habits, and the study did not take this into account. Zheng et al [44] found that the WHtR had great performance for MAFLD screening, with a sensitivity of 96% and specificity of 64%. In our study, the WHtR showed a sensitivity of 80.8% and specificity of 79.8% for MAFLD screening.

TG and FPG levels are considered 2 pivotal inducers of metabolic syndrome. TGs are produced excessively in the process of fat accumulation, and insulin resistance accelerates hepatic steatosis. The TyG index can be used as a simple alternative marker for the detection of insulin resistance in the diagnostic test combining TG and FPG levels. The prevalence and severity of MAFLD are positively correlated with the TyG index [25,45-47]. The AUC of the TyG index for predicting MAFLD was 0.835 (95% CI 4.560-9.427), which might be valuable for clinical practice.

A meta-analysis revealed that the visceral adiposity index was an independent predictor of MAFLD, which could be used to predict potential morbidity [48]. However, the predictive ability of the visceral adiposity index has not been verified. Wang et al [49] found that nonobese MAFLD patients had higher BFM and VFA values than the healthy population, and most of them had abnormal lipid metabolism. In addition, BFM and VFA were valuable for distinguishing MAFLD patients from nonobese people [49,50]. This conclusion was also confirmed in this study (BFM for the prediction of MAFLD: AUC=0.842, sensitivity=76.2%, specificity=76.9%).

This study aimed to establish a home-based model for early screening of MAFLD to promote disease self-assessment and management. Compared with previously published models that rely heavily on laboratory indicators, our model combined body composition and the WHtR to screen for MAFLD, and the body parameters that were used to build the screening model can be easily obtained using a body fat scale at home. The mobile device software can record specific values and perform calculations.

There were 2 significant advantages of our model: (1) The need for an invasive examination and medical expenditures were reduced; (2) early screening models can provide early warning signs of disease, prompting people to modify diet and exercise or seek medical treatment if necessary; (3) patient-physician interactions were enhanced.

There were also some limitations of our work. First, this study was limited by geographical factors, and regional bias existed. Second, due to ethical considerations, the results in this study cannot be confirmed by histological liver examination. Third, in some villages we went to for recruitment, we were unable to obtain a radiological diagnosis due to manpower, transportation, and other constraints. In addition, it was difficult to follow participants who underwent physical examination in different areas, and reexamination data could not be compared with previous data.

Although our study found that the new MFSI model and MRT were valuable for MAFLD prediction, disease diagnosis still requires experienced clinicians, and those with the disease or at high risk should seek timely medical attention.

## Acknowledgments

This study was funded by the National Key Research and Development Program (NO. 2018YFC2000500) and the HUAWEI (Huawei Terminal Co, Ltd) Liver Health Research Technical Cooperation Project.

## Data Availability

The data are not publicly available due to cooperative project clauses. Please contact the author to inquire if the data in this study are available for other studies.

## Conflicts of Interest

None declared.

## References

1. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2018 Jan;15(1):11-20 [FREE Full text] [doi: 10.1038/nrgastro.2017.109] [Medline: 28930295]
2. Pipitone RM, Ciccioli C, Infantino G, La Mantia C, Parisi S, Tulone A, et al. MAFLD: a multisystem disease. Ther Adv Endocrinol Metab 2023;14:20420188221145549 [FREE Full text] [doi: 10.1177/20420188221145549] [Medline: 36726391]

3. Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. J Hepatol 2020 Jul;73(1):202-209 [FREE Full text] [doi: 10.1016/j.jhep.2020.03.039] [Medline: 32278004]
4. Sun D, Jin Y, Wang T, Zheng KI, Rios RS, Zhang H, et al. MAFLD and risk of CKD. Metabolism 2021 Feb;115:154433. [doi: 10.1016/j.metabol.2020.154433] [Medline: 33212070]
5. Lin S, Huang J, Wang M, Kumar R, Liu Y, Liu S, et al. Comparison of MAFLD and NAFLD diagnostic criteria in real world. Liver Int 2020 Sep;40(9):2082-2089. [doi: 10.1111/liv.14548] [Medline: 32478487]
6. Sinn DH, Kang D, Chang Y, Ryu S, Cho SJ, Paik SW, et al. Non-alcoholic fatty liver disease and the incidence of myocardial infarction: A cohort study. J Gastroenterol Hepatol 2020 May;35(5):833-839. [doi: 10.1111/jgh.14856] [Medline: 31512278]
7. Kim H, El-Serag HB. The epidemiology of hepatocellular carcinoma in the USA. Curr Gastroenterol Rep 2019 Apr 11;21(4):17. [doi: 10.1007/s11894-019-0681-x] [Medline: 30976932]
8. Li F, Sun G, Wang Z, Wu W, Guo H, Peng L, et al. Characteristics of fecal microbiota in non-alcoholic fatty liver disease patients. Sci China Life Sci 2018 Jul;61(7):770-778. [doi: 10.1007/s11427-017-9303-9] [Medline: 29948900]
9. Allen AM, Hicks SB, Mara KC, Larson JJ, Therneau TM. The risk of incident extrahepatic cancers is higher in non-alcoholic fatty liver disease than obesity - A longitudinal cohort study. J Hepatol 2019 Dec;71(6):1229-1236 [FREE Full text] [doi: 10.1016/j.jhep.2019.08.018] [Medline: 31470068]
10. Targher G, Chonchol MB, Byrne CD. CKD and nonalcoholic fatty liver disease. Am J Kidney Dis 2014 Oct;64(4):638-652. [doi: 10.1053/j.ajkd.2014.05.019] [Medline: 25085644]
11. Targher G, Byrne CD. Non-alcoholic fatty liver disease: an emerging driving force in chronic kidney disease. Nat Rev Nephrol 2017 May;13(5):297-310. [doi: 10.1038/nrneph.2017.16] [Medline: 28218263]
12. Baratta F, Pastori D, Angelico F, Balla A, Paganini AM, Cocomello N, et al. Nonalcoholic fatty liver disease and fibrosis associated with increased risk of cardiovascular events in a prospective study. Clin Gastroenterol Hepatol 2020 Sep;18(10):2324-2331.e4. [doi: 10.1016/j.cgh.2019.12.026] [Medline: 31887443]
13. Deprince A, Haas JT, Staels B. Dysregulated lipid metabolism links NAFLD to cardiovascular disease. Mol Metab 2020 Dec;42:101092 [FREE Full text] [doi: 10.1016/j.molmet.2020.101092] [Medline: 33010471]
14. Mantovani A, Scorletti E, Mosca A, Alisi A, Byrne CD, Targher G. Complications, morbidity and mortality of nonalcoholic fatty liver disease. Metabolism 2020 Oct;111S:154170. [doi: 10.1016/j.metabol.2020.154170] [Medline: 32006558]
15. Lee SJ, Kim SU. Noninvasive monitoring of hepatic steatosis: controlled attenuation parameter and magnetic resonance imaging-proton density fat fraction in patients with nonalcoholic fatty liver disease. Expert Rev Gastroenterol Hepatol 2019 Jun;13(6):523-530. [doi: 10.1080/17474124.2019.1608820] [Medline: 31018719]
16. Nassir F. NAFLD: mechanisms, treatments, and biomarkers. Biomolecules 2022 Jun 13;12(6):1 [FREE Full text] [doi: 10.3390/biom12060824] [Medline: 35740949]
17. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, LIDO Study Group. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 2005 Jun;128(7):1898-1906. [doi: 10.1053/j.gastro.2005.03.084] [Medline: 15940625]
18. Xu L, Lu W, Li P, Shen F, Mi Y, Fan J. A comparison of hepatic steatosis index, controlled attenuation parameter and ultrasound as noninvasive diagnostic tools for steatosis in chronic hepatitis B. Dig Liver Dis 2017 Aug;49(8):910-917. [doi: 10.1016/j.dld.2017.03.013] [Medline: 28433586]
19. Karlas T, Petroff D, Sasso M, Fan J, Mi Y, de Lédinghen V, et al. Individual patient data meta-analysis of controlled attenuation parameter (CAP) technology for assessing steatosis. J Hepatol 2017 May;66(5):1022-1030. [doi: 10.1016/j.jhep.2016.12.022] [Medline: 28039099]
20. European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. Diabetologia 2016 Jun;59(6):1121-1140. [doi: 10.1007/s00125-016-3902-y] [Medline: 27053230]
21. Schattenberg JM, Loomba R. Refining noninvasive diagnostics in nonalcoholic fatty liver disease: closing the gap to detect advanced fibrosis. Hepatology 2019 Mar;69(3):934-936. [doi: 10.1002/hep.30402] [Medline: 30515858]
22. McPherson S, Hardy T, Dufour J, Petta S, Romero-Gomez M, Allison M, et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. Am J Gastroenterol 2017 May;112(5):740-751 [FREE Full text] [doi: 10.1038/ajg.2016.453] [Medline: 27725647]
23. Lee J, Vali Y, Boursier J, Spijker R, Anstee QM, Bossuyt PM, et al. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: A systematic review. Liver Int 2021 Feb;41(2):261-270 [FREE Full text] [doi: 10.1111/liv.14669] [Medline: 32946642]
24. Petta S, Wong VW, Cammà C, Hiriart J, Wong GL, Vergniol J, et al. Serial combination of non-invasive tools improves the diagnostic accuracy of severe liver fibrosis in patients with NAFLD. Aliment Pharmacol Ther 2017 Sep;46(6):617-627. [doi: 10.1111/apt.14219] [Medline: 28752524]
25. Tutunchi H, Naeini F, Mobasseri M, Ostadrahimi A. Triglyceride glucose (TyG) index and the progression of liver fibrosis: A cross-sectional study. Clin Nutr ESPEN 2021 Aug;44:483-487. [doi: 10.1016/j.clnesp.2021.04.025] [Medline: 34330512]

26. Fedchuk L, Nascimbeni F, Pais R, Charlotte F, Housset C, Ratziu V, LIDO Study Group. Performance and limitations of steatosis biomarkers in patients with nonalcoholic fatty liver disease. Aliment Pharmacol Ther 2014 Nov;40(10):1209-1222 [FREE Full text] [doi: 10.1111/apt.12963] [Medline: 25267215]
27. Lee J, Kim D, Kim HJ, Lee C, Yang JI, Kim W, et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. Dig Liver Dis 2010 Jul;42(7):503-508. [doi: 10.1016/j.dld.2009.08.002] [Medline: 19766548]
28. Lin I, Lee M, Wang C, Wu D, Chen S. Gender differences in the relationships among metabolic syndrome and various obesity-related indices with nonalcoholic fatty liver disease in a Taiwanese population. Int J Environ Res Public Health 2021 Jan 20;18(3):1 [FREE Full text] [doi: 10.3390/ijerph18030857] [Medline: 33498329]
29. Rotter I, Rył A, Grzesiak K, Szylińska A, Pawlukowska W, Lubkowska A, et al. Cross-sectional inverse associations of obesity and fat accumulation indicators with testosterone in non-diabetic aging men. Int J Environ Res Public Health 2018 Jun 08;15(6):1 [FREE Full text] [doi: 10.3390/ijerph15061207] [Medline: 29890654]
30. Verma M, Rajput M, Sahoo SS, Kaur N, Rohilla R. Correlation between the percentage of body fat and surrogate indices of obesity among adult population in rural block of Haryana. J Family Med Prim Care 2016;5(1):154-159 [FREE Full text] [doi: 10.4103/2249-4863.184642] [Medline: 27453862]
31. Jayedi A, Soltani S, Zargar MS, Khan TA, Shab-Bidar S. Central fatness and risk of all cause mortality: systematic review and dose-response meta-analysis of 72 prospective cohort studies. BMJ 2020 Sep 23;370:m3324 [FREE Full text] [doi: 10.1136/bmj.m3324] [Medline: 32967840]
32. Cai J, Lin C, Lai S, Liu Y, Liang M, Qin Y, et al. Waist-to-height ratio, an optimal anthropometric indicator for metabolic dysfunction associated fatty liver disease in the Western Chinese male population. Lipids Health Dis 2021 Oct 27;20(1):145 [FREE Full text] [doi: 10.1186/s12944-021-01568-9] [Medline: 34706716]
33. Byrne CD, Patel J, Scorletti E, Targher G. Tests for diagnosing and monitoring non-alcoholic fatty liver disease in adults. BMJ 2018 Jul 12;362:k2734. [doi: 10.1136/bmj.k2734] [Medline: 30002017]
34. Hagström H, Nasr P, Ekstedt M, Stål P, Hultcrantz R, Kechagias S. Accuracy of noninvasive scoring systems in assessing risk of death and liver-related endpoints in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2019 May;17(6):1148-1156.e4. [doi: 10.1016/j.cgh.2018.11.030] [Medline: 30471458]
35. Sasso M, Beaugrand M, de Ledinghen V, Douvin C, Marcellin P, Poupon R, et al. Controlled attenuation parameter (CAP): a novel VCTE™ guided ultrasonic attenuation measurement for the evaluation of hepatic steatosis: preliminary study and validation in a cohort of patients with chronic liver disease from various causes. Ultrasound Med Biol 2010 Nov;36(11):1825-1835. [doi: 10.1016/j.ultrasmedbio.2010.07.005] [Medline: 20870345]
36. Mikolasevic I, Milic S, Orlic L, Stimac D, Franjic N, Targher G. Factors associated with significant liver steatosis and fibrosis as assessed by transient elastography in patients with one or more components of the metabolic syndrome. J Diabetes Complications 2016;30(7):1347-1353. [doi: 10.1016/j.jdiacomp.2016.05.014] [Medline: 27324703]
37. Almeida NS, Rocha R, Cotrim HP, Daltro C. Anthropometric indicators of visceral adiposity as predictors of non-alcoholic fatty liver disease: A review. World J Hepatol 2018 Oct 27;10(10):695-701 [FREE Full text] [doi: 10.4254/wjh.v10.i10.695] [Medline: 30386462]
38. Motamed N, Rabiee B, Hemasi GR, Ajdarkosh H, Khonsari MR, Maadi M, et al. Body roundness index and waist-to-height ratio are strongly associated with non-alcoholic fatty liver disease: a population-based study. Hepat Mon 2016 Sep;16(9):e39575 [FREE Full text] [doi: 10.5812/hepatmon.39575] [Medline: 27822266]
39. Wu Y, Kumar R, Wang M, Singh M, Huang J, Zhu Y, et al. Validation of conventional non-invasive fibrosis scoring systems in patients with metabolic associated fatty liver disease. World J Gastroenterol 2021 Sep 14;27(34):5753-5763 [FREE Full text] [doi: 10.3748/wjg.v27.i34.5753] [Medline: 34629799]
40. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology 2007 Apr;45(4):846-854. [doi: 10.1002/hep.21496] [Medline: 17393509]
41. Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ, Nash Clinical Research Network. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2009 Oct;7(10):1104-1112 [FREE Full text] [doi: 10.1016/j.cgh.2009.05.033] [Medline: 19523535]
42. Harrison SA, Oliver D, Arnold HL, Gogia S, Neuschwander-Tetri BA. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. Gut 2008 Oct;57(10):1441-1447. [doi: 10.1136/gut.2007.146019] [Medline: 18390575]
43. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol 2006 Nov 02;6:33 [FREE Full text] [doi: 10.1186/1471-230X-6-33] [Medline: 17081293]
44. Zheng R, Chen Z, Chen J, Lu Y, Chen J. Role of body mass index, waist-to-height and waist-to-hip ratio in prediction of nonalcoholic fatty liver disease. Gastroenterol Res Pract 2012;2012:362147 [FREE Full text] [doi: 10.1155/2012/362147] [Medline: 22701476]
45. Farrell GC. Signalling links in the liver: knitting SOCS with fat and inflammation. J Hepatol 2005 Jul;43(1):193-196. [doi: 10.1016/j.jhep.2005.04.004] [Medline: 15913829]

46. Zhang S, Du T, Li M, Jia J, Lu H, Lin X, et al. Triglyceride glucose-body mass index is effective in identifying nonalcoholic fatty liver disease in nonobese subjects. Medicine (Baltimore) 2017 Jun;96(22):e7041 [FREE Full text] [doi: 10.1097/MD.0000000000007041] [Medline: 28562560]
47. Zhang S, Du T, Zhang J, Lu H, Lin X, Xie J, et al. The triglyceride and glucose index (TyG) is an effective biomarker to identify nonalcoholic fatty liver disease. Lipids Health Dis 2017 Jan 19;16(1):15 [FREE Full text] [doi: 10.1186/s12944-017-0409-6] [Medline: 28103934]
48. Yi X, Zhu S, Zhu L. Diagnostic accuracy of the visceral adiposity index in patients with metabolic-associated fatty liver disease: a meta-analysis. Lipids Health Dis 2022 Mar 06;21(1):28 [FREE Full text] [doi: 10.1186/s12944-022-01636-8] [Medline: 35249545]
49. Wang YJ, Cheng HR, Zhou WH. Correlation of body fat composition and metabolic indicators with metabolic-associated fatty liver disease in a non-obese population. Chinese General Practice 2023;26(6):672-680. [doi: 10.12114/j.issn.1007-9572.2022.0573]
50. Byrne CD, Targher G. Ectopic fat, insulin resistance, and nonalcoholic fatty liver disease: implications for cardiovascular disease. Arterioscler Thromb Vasc Biol 2014 Jun;34(6):1155-1161. [doi: 10.1161/ATVBAHA.114.303034] [Medline: 24743428]

## Abbreviations

**ALT:** alanine aminotransferase
**APRI:** aspartate aminotransferase-to-platelet ratio index
**AST:** aspartate aminotransferase
**AUC:** area under the curve
**BFM:** body fat mass
**BMR:** basal metabolic rate
**CAP:** controlled attenuation parameter
**FIB-4:** fibrosis-4
**FLD:** fatty liver disease
**FLI:** fatty liver index
**FPG:** fasting plasma glucose
**GGT:** glutamyl transpeptidase
**GPR:** glutamyl transpeptidase-to-platelet ratio index
**HDL-c:** high-density lipoprotein cholesterol)
**HSI:** hepatic steatosis index
**LDL-c:** low-density lipoprotein cholesterol
**MAFLD:** metabolically associated fatty liver disease
**MFSI:** MAFLD screening index
**MRT:** MAFLD Rating Table
**NAFLD:** nonalcoholic fatty liver disease
**NFS:** nonalcoholic fatty liver disease fibrosis score
**OR:** odds ratio
**PLT:** platelet
**ROC:** receiver operating characteristic
**TBW:** total body water
**TC:** total cholesterol
**TG:** triglyceride
**TyG:** triglyceride glucose
**VFA:** visceral fat area
**WHtR:** waist-height ratio

*Edited by A Mavragani; submitted 03.01.24; peer-reviewed by Y Zhang, H Yu; comments to author 23.05.24; revised version received 03.06.24; accepted 03.07.24; published 22.08.24.*

*Please cite as:*
*Ni J, Huang Y, Xiang Q, Zheng Q, Xu X, Qin Z, Sheng G, Li L*
*Establishment and Evaluation of a Noninvasive Metabolism-Related Fatty Liver Screening and Dynamic Monitoring Model: Cross-Sectional Study*
*Interact J Med Res 2024;13:e56035*
*URL: https://www.i-jmr.org/2024/1/e56035*
*doi:10.2196/56035*
*PMID:39172506*

Original Paper

# Perception of Medication Safety–Related Behaviors Among Different Age Groups: Web-Based Cross-Sectional Study

Yan Lang[1], PhD; Kay-Yut Chen[2], PhD; Yuan Zhou[3], PhD; Ludmila Kosmari[4], RN; Kathryn Daniel[4†], PhD, RN; Ayse Gurses[5], MS, MPH, PhD; Richard Young[6], MD; Alicia Arbaje[7], MPH, MD, PhD; Yan Xiao[4], PhD

[1]Department of Business, State University of New York at Oneonta, Oneonta, NY, United States

[2]College of Business, University of Texas at Arlington, Arlington, TX, United States

[3]Department of Industrial, Manufacturing, and Systems Engineering, University of Texas at Arlington, Arlington, TX, United States

[4]College of Nursing and Health Innovation, University of Texas at Arlington, Arlington, TX, United States

[5]School of Medicine, Johns Hopkins University, Baltimore, MD, United States

[6]Family Medicine Residency Program, The John Peter Smith (JPS) Health Network, Fort Worth, TX, United States

[7]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[†]deceased

**Corresponding Author:**
Yan Lang, PhD
Department of Business
State University of New York at Oneonta
108 Ravine Pkwy
Oneonta, NY, 13820
United States
Phone: 1 607 436 3251
Email: yan.lang@oneonta.edu

## Abstract

**Background:** Previous research and safety advocacy groups have proposed various behaviors for older adults to actively engage in medication safety. However, little is known about how older adults perceive the importance and reasonableness of these behaviors in ambulatory settings.

**Objective:** This study aimed to assess older adults' perceptions of the importance and reasonableness of 8 medication safety behaviors in ambulatory settings and compare their responses with those of younger adults.

**Methods:** We conducted a survey of 1222 adults in the United States using crowdsourcing to evaluate patient behaviors that may enhance medication safety in community settings. A total of 8 safety behaviors were identified based on the literature, such as bringing medications to office visits, confirming medications at home, managing medication refills, using patient portals, organizing medications, checking medications, getting help, and knowing medications. Respondents were asked about their perception of the importance and reasonableness of these behaviors on a 5-point Likert rating scale in the context of collaboration with primary care providers. We assessed the relative ranking of behaviors in terms of importance and reasonableness and examined the association between these dimensions across age groups using statistical tests.

**Results:** Of 1222 adult participants, 125 (10.2%) were aged 65 years or older. Most participants were White, college-educated, and had chronic conditions. Older adults rated all 8 behaviors significantly higher in both importance and reasonableness than did younger adults ($P$<.001 for combined behaviors). Confirming medications ranked highest in importance (mean score=3.78) for both age groups while knowing medications ranked highest in reasonableness (mean score=3.68). Using patient portals was ranked lowest in importance (mean score=3.53) and reasonableness (mean score=3.49). There was a significant correlation between the perceived importance and reasonableness of the identified behaviors, with coefficients ranging from 0.436 to 0.543 (all $P$<.001).

**Conclusions:** Older adults perceived the identified safety behaviors as more important and reasonable than younger adults. However, both age groups considered a behavior highly recommended by professionals as the least important and reasonable. Patient engagement strategies, common and specific to age groups, should be considered to improve medication safety in ambulatory settings.

*(Interact J Med Res 2024;13:e58635)* doi:10.2196/58635

**KEYWORDS**

medication safety; patient engagement; aged adults; survey; Amazon Mechanical Turk; medication; engagement; older adults; elderly; safety; United States; USA; crowdsourcing; community; patient portal; primary care; medications; safety behavior; younger adults; age; correlation; statistical test

## Introduction

Engagement of older adult patients has been recognized as key to health outcomes including safety [1]. While specific skills and attitudes related to patient engagement have been identified and measured [2,3], there is a lack of clarity regarding the specific roles and responsibilities expected of patients in community settings, where patients and families are responsible for the medication use process. Health care organizations often set implicit expectations regarding the roles and responsibilities of patients and their families in collaborative activities such as planning, implementation, and discourse about their health [4].

The purpose of our study was to use a crowdsourcing approach to investigate individuals' perceptions of the importance and reasonableness of medication safety behaviors across various age groups. We chose patient portal use as a reference for patient engagement behaviors due to extensive efforts by health care organizations and regulators to encourage this behavior. A survey conducted in 2020 showed that more than half of individuals nationwide were offered access to patient portals, with nearly 40% accessing their records [5]. Patients' perspectives on the importance and reasonableness of using portals are important to understand in order to devise interventions to encourage the behavior, such as patients' interest, willingness, and ability [6-8], especially among older adults [9].

Health care professionals are encouraged to guide patients and their families to actively participate in their care by adopting safety behaviors, although little is known about older adults' perspectives on these behaviors and roles within the collaborative process to improve patient engagement in medication safety. By understanding the perspectives of laypeople, we can identify the gaps in engaging patients and family members in medication safety improvements and design interventions that can better meet their needs. Furthermore, understanding the relation between importance and reasonableness in perceiving medication safety behaviors is crucial for health care professionals to tailor their interventions and communication strategies effectively, particularly for older adults, to promote safer medication practices.

The aim of this study is to assess how older adults perceive the importance and reasonableness of 8 medication safety behaviors in ambulatory settings and to compare their responses with those of younger adults.

## Methods

### Study Design

This cross-sectional study was conducted using a role-playing survey to assess the importance and reasonableness of medication safety behaviors.

### Setting

The study was conducted using Amazon Mechanical Turk (MTurk), a crowdsourcing platform, from October to December 2022. Participants completed the survey online. This approach allowed us to efficiently gather data from a large group of participants [10].

### Participants

Our study limited participants to US adults (≥18 years old) who had established a strong reputation on MTurk, defined as completed 100+ tasks with at least 95% approval ratings [11]. This choice leverages the acknowledged representativeness of US MTurk samples for diverse psychological dimensions [12] while ensuring engaged and reliable participation, as users with good reputations are generally more motivated and provide accurate data [13]. We used the Software Platform for Human Interaction Experiments (SoPHIE; SoPHIELabs) to administer the surveys. Participants were screened for eligibility through SoPHIE and were required to read and sign a consent form before participating. Qualified participants were given an online consent form, where they expressed their voluntary agreement to participate by clicking on the "Continue" button on their computer screen.

### Assessments

Participants were asked to envision themselves as older adults, retired individuals living alone with multiple health conditions (detailed instructions in Figure S1 in Multimedia Appendix 1). To identify safety behaviors in managing medication use in ambulatory settings, we reviewed literature and recommendations from safety organizations to represent professionals' views on what patients should do to contribute to medication safety. For example, 1 study targeted behaviors associated with an office visit for patient engagement, including writing out a list of medications or bringing medications to visit [14]. In our survey, we defined the "importance" of a behavior as the extent to which all patients and families should adopt it for medication safety. "Reasonableness" was judged based on the assumption that following a treatment regimen makes sense if it leads to better health outcomes [15].

### Study Survey

This study used a carefully developed survey instrument. Initial pilot studies, with 14 closed-ended questions, assessed medication safety behaviors. Based on feedback from

participants and experts, the survey was refined and consolidated for clarity and focus, resulting in a final 8-item instrument (Table 1). Throughout pilot testing, we iteratively evaluated the content validity of the questions against existing literature and organizational recommendations in medication safety. While patient involvement in developing the criteria and indicators was not direct, they were informed by a comprehensive review of relevant literature, safety advocacy group recommendations, and expert consensus in the field. Safe self-administration of medication heavily relies on patients' knowledge about their treatments [16], their purpose, proper usage instructions, identifying and reporting adverse effects, obtaining refills, and effectively communicating any issues related to their prescribed medications with their health care provider. While the 8 behaviors specified in this study are not exhaustive, they encompass these crucial components and are presented for ease of understanding and application.

**Table 1.** Targeted patient behaviors in medication safety used in the survey.

| Medication safety behaviors | Examples provided to participants | Justifications and references |
|---|---|---|
| **Bring medications** | | |
| Patients are expected to bring all medications and all relevant health-related documents to their health care provider office visits. | Collect all medicine bottles, including those over the counter such as Tylenol and vitamins, and bring them with you to the health care provider's office. Make sure to also bring documents such as medication lists and blood sugar and blood pressure logs (if asked to keep one). | • The FDA[a] recommends keeping a list of all medications (prescribed and over the counter) and bringing it to all doctors' appointments [17].<br>• Although often encouraged by primary care providers [18,19].<br>• Only about 20%-40% of the patients bring in their medications or medication list.<br>• Medication reviews lack standardization which can result in increased mortality, morbidity, and poor patient outcomes [20,21]. |
| **Confirm medications** | | |
| Patients will verify any changes in their medications after each provider's visit. | You have been taking 20 mg of Simvastatin every day for cholesterol for a long time. In the last visit, your cholesterol level has decreased. Your provider reduced the medication dose to 10 mg. You make this change on your personal medication list. | • FDA recommends to verify the medication list at least once a year or any time there is a change [17].<br>• Medication discrepancies are very common among patients with chronic conditions, especially those who require frequent hospitalizations or see numerous providers [22,23]. |
| **Refill system** | | |
| Patients will establish a refill system. | Your provider advised you to call the pharmacy when you are about to run out of refills, and not their office. The pharmacy will contact the provider's office for refill prescriptions. Using two 7-day pill boxes allows you to know 2 weeks in advance when a medicine will run out. | • Current recommendations are to address all refill needs during the provider's visit and send all prescriptions ideally to one pharmacy only in order to prevent gaps in medication therapies in chronic care due to disruptions and lapses in obtaining refills timely. Innovative systems use technological advances; however, older patients and individuals who speak English as a second language are less likely to use technology to refill medications [24-28]. |
| **Use portals** | | |
| Using patient health care portals. | Your provider's office sent you a link for creating an account to access the patient portal website. After you sign up, you can use the portal to communicate with your provider and access your health information. | • Patient portals were intended to improve the communication between the health care team and patients. They allow patients to be actively involved in their care, access their medical records, verify for accuracy, report concerns, and seek medical advice or medication refill [6,8,29-31]. |
| **Organize medications** | | |
| Using pill dispensers and other organizer tools. | Pill boxes are effective tools to remind you when and what medicines to take. You may also set reminders on your phone. To-go boxes are convenient to carry in your bag or purse when you are out, running errands. Charts, calendars, and electronic pill boxes are other ways of organizing medications. | • The importance of having a system to organize medications was extensively studied. A list of memory tips and reminder systems (such as daily pillbox, calendar, or chart) to help organize scheduled prescriptions are available on various online resources [32-35]. |
| **Check medications** | | |
| Verifying medications for duplicates and expired medications. | The mail-order pharmacy sends you your refills automatically, so they always arrive before you run out of it. You know how to check the medicines against your list, as the color of the pills and names of medications (eg, generic vs brand name) may change from time to time, and you do not want to take duplicate medicines. You also dispose of expired medicines, so you do not accidentally take them. | • The FDA and NIH[b] recommend that patients check all medications for expiration dates. Ingesting expired medications may pose significant health hazards [17,36]. |
| **Medication awareness** | | |

| Medication safety behaviors | Examples provided to participants | Justifications and references |
|---|---|---|
| Accessing resources pertaining to medication-related issues. | You went to the pharmacy to pick up a prescription, but they did not have it. To clarify the situation, you call your provider's office to inquire if you are still supposed to take the medication and verify the correct pharmacy on file.<br>In another situation, you may need to call the pharmacist or the provider to find out what you need to do if you accidentally doubled your heart medicine. | • The World Health Organization recommends that patients learn to identify and report any issues or side effects pertaining to taking medication [17,36]. |
| **Know medications** | | |
| Have basic knowledge about medications. | When you look at your medicine bottle, you are able to locate the medicine name, dose, when and how to take this medicine, how many refills are left, expiration date, and telephone number to call if you have questions about this medicine. For instance, you are prescribed to take a round white pill twice a day for high blood pressure. You wrote the name ("metoprolol") on the medication list. You know to take 1 pill in the morning and 1 pill in the evening. You also know that you should take the pill with food. Symptoms to watch for are lightheadedness or very slow heartbeats. | • Patients are advised to read carefully all information provided with the medications such as package inserts and pharmacy instructions. As many as an estimated 87% of patients do not read these instructions. Health illiteracy continues to be a challenge. Patients who are younger and have a higher formal education are more likely to have an adequate knowledge of medications [35,37-39]. |

[a]FDA: US Food and Drug Administration.

[b]NIH: National Institutes of Health.

## Data Sources

Participants were asked to rate their perceptions of these behaviors in terms of importance and reasonableness on a 5-point Likert scale, with response categories of 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), and 5 (strongly agree). After completing the survey, participants were asked to provide their demographic information (age, sex, race, ethnicity, education, and income) and number of chronic medical conditions.

## Study Size

The study size was determined by targeting a minimum of 1000 participants to ensure sufficient power to detect differences between age groups. This target was based on previous studies in the field that used similar methodologies and sample sizes to achieve robust statistical power and generalizability of findings [40,41]. We received 1222 completed surveys, achieving a completion rate of 94.5%, of the total 1293 attempts.

## Data Analysis

Statistical analyses were performed using Stata software (version 17; StataCorp). Differences in importance and reasonableness between 2 age groups (younger than 65 years and 65 years or older) were assessed by the Wilcoxon rank-sum tests. The associations between importance and reasonableness were assessed by the Pearson correlation coefficient. Disagreement was assessed by differences in ratings of reasonableness and importance scores for the same behavior by the same participant, with serious disagreement defined as a difference of 3 or more points. For ordered logistic regression models, the outcome variables were the perceived importance and reasonableness of the 8 medication safety behaviors, while the independent variables included age group, sex, ethnicity, education, income, and the number of chronic medical conditions. Each model used only 1 independent variable at a time to assess its individual impact on the outcome. Subgroup analyses with ordered logistic regression based on sex, ethnicity, education, income, and chronic medical conditions were conducted to understand how various factors influenced age-related perceptions of the 8 health behaviors.

## Ethical Considerations

### *Human Participant Ethics Review Approvals or Exemptions*

This study was approved by the institutional review board (IRB) at the University of Texas at Arlington (protocol 2022-0581). The research involved human participants and adhered to appropriate ethical review and approvals as per institutional guidelines.

### *Informed Consent*

Participants were given an online consent form, detailing the study's purpose, procedures, potential risks, and benefits. They expressed their voluntary agreement to participate by clicking on the "Continue" button on their computer screen. Participants were informed of their ability to opt out of the study at any time without any consequences.

### *Privacy and Confidentiality*

All data collected were anonymized to protect the privacy and confidentiality of participants. No personally identifiable information was collected. Data were stored securely on password-protected servers, and only the research team had access to the anonymized data set.

### *Compensation Details*

Participants were compensated for their time and effort. Each participant received US $0.25 for completing the survey, which is a standard compensation rate for similar studies conducted on MTurk. This compensation was designed to be fair and transparent, ensuring that participants were adequately reimbursed for their contribution to the research.

## Results

We received 1222 completed surveys (completion rate of 94.5%) from a total of 1293 attempts in December 2022. Participants took an average of 3 minutes to complete the survey. The majority of the participants were younger than 65 years, White, held a bachelor's degree, reported an income range of US $40,000-80,000, and had 1 or more chronic conditions (Table 2). Detailed comparisons on sex, race, education, income, and number of chronic conditions are reported in the Multimedia Appendix 1.

Across age groups, "confirming medications" was rated as the most important behavior while "knowing medications" was rated as the most reasonable behavior (Figure 1). In contrast, the behavior of using portals received the lowest scores for both importance and reasonableness. The perceived importance and reasonableness of each behavior were positively correlated ($P$<.001; correlations in Table S1 in Multimedia Appendix 1). Serious disagreements between importance and reasonableness for the 8 behaviors were between 6% and 7.3% among the participants (Table S2 in Multimedia Appendix 1). In addition, older adult participants reported higher importance and lower reasonableness ratings across the 8 identified behaviors (Multimedia Appendix 2).

Among older adult participants, the behavior of confirming medication was scored highest in importance, whereas the younger age participants scored the behavior of bringing medications as highest in importance (Multimedia Appendix 3). Furthermore, among the older adult participants, the behaviors of getting help ($P$=.007) and knowing medications ($P$=.03) were rated as the second and third most important, respectively, and were found to be significantly higher than those rated by younger adult participants.

The behavior of using portals scored lowest in both reasonableness and importance for both age groups. In addition, there was no significant difference in the scoring of this behavior between the 2 age groups. Out of the total, 3 of the remaining behaviors were scored significantly higher by older adult participants than younger adult participants (Multimedia Appendix 3).

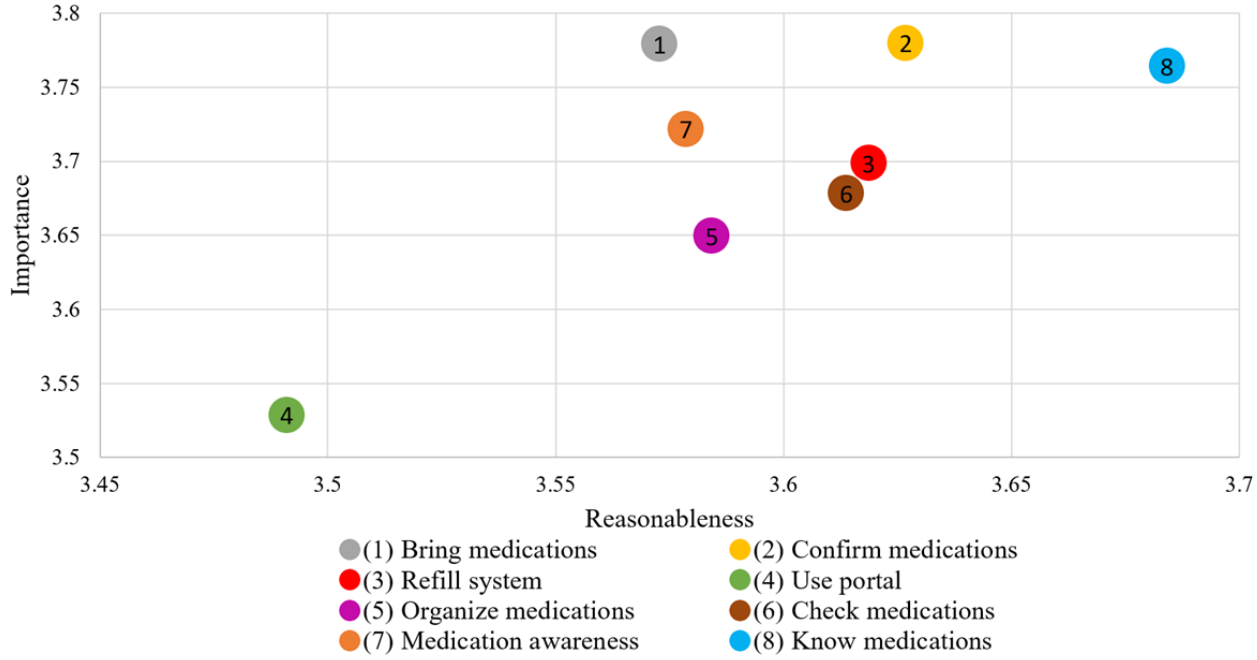
Across different age groups, we observed a general negative association between the perceived importance and reasonableness of these behaviors with male, Hispanic or Latino ethnicity, and education levels. Conversely, higher income was associated with a positive perception of importance. In addition, chronic medical conditions were linked to a negative perception of reasonableness. A detailed breakdown of these findings by individual behavior is shown in the Multimedia Appendix 1.

**Table 2.** Demographics of study participants (n=1222).

| Characteristic | <65 years | ≥65 years | Proportion test across age groups, *P* value |
|---|---|---|---|
| Age (years), mean (SD) | 36.77 (10.73) | 68.70 (2.95) | —[a] |
| Female sex, n (%) | 497 (45.31) | 64 (51.2) | .21 |
| **Race, n (%)** | | | |
| White | 934 (85.1) | 83 (66.4) | <.001 |
| Non-White | 163 (14.9) | 42 (33.6) | <.001 |
| Hispanic or Latino ethnicity, n (%) | 327 (29.8) | 40 (32) | .61 |
| **Education, n (%)** | | | |
| Less than bachelor's degree | 131 (11.9) | 42 (33.6) | <.001 |
| Bachelor's degree | 679 (61.9) | 45 (36) | <.001 |
| Graduate degree | 287 (26.2) | 38 (30.4) | .31 |
| **Annual household income (US $), n (%)** | | | |
| <40,000 | 300 (27.3) | 44 (35.2) | .07 |
| 40,000-80,000 | 659 (60.1) | 70 (56) | .38 |
| >80,000 | 138 (12.6) | 11 (8.8) | .22 |
| **Chronic medical conditions, n (%)** | | | |
| 0 | 418 (38.1) | 24 (19.2) | <.001 |
| ≥1 | 679 (61.9) | 101 (80.8) | <.001 |

[a]—: not applicable.

**Figure 1.** Perceptions of behaviors in terms of importance and reasonableness (n=1222). Range scale was used.




## Discussion

Our study found that older adults perceive a higher importance and reasonableness of medication safety behaviors compared with younger adults. Specifically, confirming medications and knowing medications were rated as the most important and reasonable behaviors across age groups while using patient portals was perceived as the least important and reasonable.

Previous research and safety advocacy organizations have suggested behaviors that patients can adopt to improve medication safety in ambulatory settings. Figure 1 provides an overview of participants' perceptions of each behavior in terms of importance and reasonableness. Despite advocacy from policy makers and professional organizations for patient portal usage [6,42], the results show that it is perceived as the least reasonable and important behavior across age groups. This could be due to a perceived disconnection between portals and medication-related safety, although our study did not directly explore this link. It is important to note that the survey question about patient portal usage did not explicitly link its use to performing other medication safety behaviors, which may have influenced participants' perceptions. Nonetheless, the discrepancy highlights the importance of understanding patients' perspectives in shaping policy and clinical practices.

Conversely, the behaviors of confirming medications and knowing medications received the highest scores in terms of both importance and reasonableness. These findings emphasize the significance of these behaviors in patient engagement and medication safety efforts. Health care providers should recognize the value placed by patients on these aspects of medication management and incorporate discussions and interventions related to these behaviors into clinical practice.

Our analysis also revealed a strong correlation between reasonableness and importance for all behaviors among participants. This suggests that intervention strategies may consider targeting efforts to explain the importance and values of these behaviors. This aligns with a cost-benefit thinking approach, where something is considered reasonable when its benefits (ie, importance) are perceived as higher than the efforts required [43]. By emphasizing the benefits of engaging in medication safety practices, health care providers can encourage patients to adopt these behaviors more effectively.

Our results suggested that older adults may be more cognizant of and experienced with health issues, making them more willing to expend effort to carry out the identified behaviors. Further research is needed to understand the reasons behind this observed difference in scores between the 2 age groups.

Our study has several limitations. While role-playing experiments are widely used in marketing science and health care [10,44], the online platform used in the study, MTurk, may introduce biases in the sample population [45]. Consequently, the perspectives of our participants may not fully represent the broader demographic, particularly older adults, in terms of sex, race, education level, and chronic conditions. Furthermore, the underrepresentation of participants older than 65 years (10% vs 17% in the US population) could limit the generalizability of our findings. However, we mitigated this by analyzing participants older than 65 years separately from those younger than 65 years, allowing us to explore differences in perceptions of medication safety behaviors between these distinct age groups.

In addition, it is important to recognize the complexity inherent in assessing perceptions of importance, particularly across diverse demographic groups. While our study aimed to capture patient perspectives on medication safety behaviors, the construct of importance is multifaceted and may be influenced by individual experiences, beliefs, and priorities. This highlights the potential for bias in responses when using a single age frame,

particularly for younger participants who might underestimate the capabilities and perspectives of older adults.

Furthermore, the survey question about patient portal usage did not explicitly link its use to performing other medication safety behaviors, which may have influenced participants' perceptions. The lack of context regarding the comprehensive use of patient portals could have impacted the ratings given by participants, thus presenting a limitation in accurately assessing the perceived importance and reasonableness of using patient portals.

Future research should consider using alternative sampling methods, such as stratified sampling or oversampling of underrepresented groups, to ensure greater representativeness in the sample. Conducting similar studies with a purposive sampling strategy could provide a more comprehensive understanding of patient perceptions of medication safety behaviors across diverse demographic groups, including older adults with multiple chronic conditions. Given the study's limited scope, qualitative components were not integrated into the survey. Subsequent studies should encompass qualitative interviews with a spectrum of individuals, both laypersons and professionals. Furthermore, careful consideration should be given to the age-related framing of survey questions. Using age-specific framing tailored to different groups, such as presenting scenarios relevant to their age experiences, could minimize bias and provide more accurate insights into participant perceptions and behaviors across various age ranges, ultimately leading to findings with greater validity and generalizability.

## Conclusion

Our study found that older adults perceive higher importance of a set of safety medication behaviors and see these behaviors as more reasonable to perform than younger adults. Using portals is generally perceived lower in importance and reasonableness by patients in ambulatory settings when compared with other medication safety behaviors, such as bringing medications to clinic visits. Future studies should explore additional factors influencing patient engagement in medication safety behaviors including social determinants of health. Longitudinal studies are needed to understand how improvement efforts can take advantage of patient perspectives on medication safety to design interventions to encourage the adoption of specific behaviors.

## Acknowledgments

This project was supported by a grant (R18HS027277) from the Agency for Healthcare Research and Quality (AHRQ). The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ. We benefited from the contribution of the PROMIS (Partnership for Resilience in Medication Safety) Lab members, including Noah Hendrix, Kimbley Fulda, Anna Espinoza, Kathleen Sutcliffe, and Samantha Pitts.

## Data Availability

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

## Conflicts of Interest

As per the University of Texas at Arlington policy, the following statement is included. K-YC has a potential research conflict of interest due to a financial interest with companies Hewlett-Packard Enterprise, Boostr, and DecisionNext. A management plan has been created to preserve objectivity in research in accordance with the University of Texas at Arlington policy. All other authors had neither competing financial interests nor other potential conflicts of interest.

Multimedia Appendix 1
Mturk Experimental Instructions.
[DOCX File , 304 KB - ijmr_v13i1e58635_app1.docx ]

Multimedia Appendix 2
Perceptions of combined behaviors in terms of importance and reasonableness among age groups (n=1097 for age <65 years and n=125 for age ≥65 years).
[DOCX File , 73 KB - ijmr_v13i1e58635_app2.docx ]

Multimedia Appendix 3
Perceptions of behaviors in terms of importance and reasonableness among age groups (n=1097 for age <65 years and n=125 for age ≥65 years). Wilcoxon rank-sum test $P<.001$.
[DOCX File , 177 KB - ijmr_v13i1e58635_app3.docx ]

## References

1. Sharma AE, Rivadeneira NA, Barr-Walker J, Stern RJ, Johnson AK, Sarkar U. Patient engagement in health care safety: an overview of mixed-quality evidence. Health Aff (Millwood) 2018;37(11):1813-1820 [FREE Full text] [doi: 10.1377/hlthaff.2018.0716] [Medline: 30395509]
2. Hibbard JH, Collins PA, Mahoney E, Baker LH. The development and testing of a measure assessing clinician beliefs about patient self-management. Health Expect 2010;13(1):65-72 [FREE Full text] [doi: 10.1111/j.1369-7625.2009.00571.x] [Medline: 19906211]
3. Kim JM, Suarez-Cuervo C, Berger Z, Lee J, Gayleard J, Rosenberg C, et al. Evaluation of patient and family engagement strategies to improve medication safety. Patient 2018;11(2):193-206. [doi: 10.1007/s40271-017-0270-8] [Medline: 28795338]
4. Batalden M, Batalden P, Margolis P, Seid M, Armstrong G, Opipari-Arrigan L, et al. Coproduction of healthcare service. BMJ Qual Saf 2016;25(7):509-517 [FREE Full text] [doi: 10.1136/bmjqs-2015-004315] [Medline: 26376674]
5. Health Information National Trends Survey 5 (HINTS 5), Cycle 4 Methodology Report.: Westat URL: https://hints.cancer.gov/docs/methodologyreports/HINTS5_Cycle4_MethodologyReport.pdf [accessed 2023-06-18]
6. Neuner J, Fedders M, Caravella M, Bradford L, Schapira M. Meaningful use and the patient portal: patient enrollment, use, and satisfaction with patient portals at a later-adopting center. Am J Med Qual 2015;30(2):105-113 [FREE Full text] [doi: 10.1177/1062860614523488] [Medline: 24563085]
7. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
8. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
9. Zoorob D, Hasbini Y, Chen K, Wangia-Anderson V, Moussa H, Miller B, et al. Ageism in healthcare technology: the older patients' aspirations for improved online accessibility. JAMIA Open 2022;5(3):ooac061 [FREE Full text] [doi: 10.1093/jamiaopen/ooac061] [Medline: 35855421]
10. Chen K, Lang Y, Zhou Y, Kosmari L, Daniel K, Gurses A, et al. Assessing interventions on crowdsourcing platforms to nudge patients for engagement behaviors in primary care settings: randomized controlled trial. J Med Internet Res 2023;25:e41431 [FREE Full text] [doi: 10.2196/41431] [Medline: 37440308]
11. Chen K, Wang J, Lang Y. Coping with digital extortion: an experimental study of benefit appeals and normative appeals. Manag Sci 2022;68(7):5269-5286. [doi: 10.1287/mnsc.2021.4154]
12. O'Connell K, Berluti K, Rhoads S, Marsh A. Reduced social distancing early in the COVID-19 pandemic is associated with antisocial behaviors in an online United States sample. PLoS One 2021;16(1):e0244974 [FREE Full text] [doi: 10.1371/journal.pone.0244974] [Medline: 33412567]
13. Arch JJ, Carr AL. Using mechanical turk for research on cancer survivors. Psychooncology 2017;26(10):1593-1603. [doi: 10.1002/pon.4173] [Medline: 27283906]
14. Tennstedt SL. Empowering older patients to communicate more effectively in the medical encounter. Clin Geriatr Med 2000;16(1):61-70, ix. [doi: 10.1016/s0749-0690(05)70008-2] [Medline: 10723618]
15. Kini V, Ho PM. Interventions to improve medication adherence: a review. JAMA 2018;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
16. Osterberg L, Blaschke T. Adherence to Medication. N Engl J Med 2005;353(5):487-497. [doi: 10.1056/nejmra050100]
17. 4 Medication safety tips for older adults. FDA. URL: https://www.fda.gov/consumers/consumer-updates/4-medication-safety-tips-older-adults [accessed 2023-05-30]
18. Jäger C, Steinhaeuser J, Freund T, Szecsenyi J, Goetz K. Medication lists and brown bag reviews: potential positive and negative impacts on patients beliefs about their medicine. Biomed Res Int 2015;2015:874067 [FREE Full text] [doi: 10.1155/2015/874067] [Medline: 26539533]
19. Weiss BD, Brega AG, LeBlanc WG, Mabachi NM, Barnard J, Albright K, et al. Improving the effectiveness of medication review: guidance from the health literacy universal precautions toolkit. J Am Board Fam Med 2016;29(1):18-23 [FREE Full text] [doi: 10.3122/jabfm.2016.01.150163] [Medline: 26769873]
20. Murtha E, Elder B, Faragher M. Brown bag medication review: using AHRQ's brown bag medication tool. J Nurs Care Qual 2020;35(1):58-62. [doi: 10.1097/NCQ.0000000000000399] [Medline: 30889082]
21. Twigg G, David T, Taylor J. An improved comprehensive medication review process to assess healthcare outcomes in a rural independent community pharmacy. Pharmacy (Basel) 2019;7(2):66 [FREE Full text] [doi: 10.3390/pharmacy7020066] [Medline: 31212922]
22. Brody AA, Gibson B, Tresner-Kirsch D, Kramer H, Thraen I, Coarr ME, et al. High prevalence of medication discrepancies between home health referrals and centers for medicare and medicaid services home health certification and plan of care and their potential to affect safety of vulnerable elderly adults. J Am Geriatr Soc 2016;64(11):e166-e170 [FREE Full text] [doi: 10.1111/jgs.14457] [Medline: 27673753]
23. Caleres G, Modig S, Midlöv P, Chalmers J, Bondesson Å. Medication discrepancies in discharge summaries and associated risk factors for elderly patients with many drugs. Drugs Real World Outcomes 2020;7(1):53-62 [FREE Full text] [doi: 10.1007/s40801-019-00176-5] [Medline: 31834621]

24. Belyeu BM, Klein JW, Reisch LM, Peacock S, Oster NV, Elmore JG, et al. Patients' perceptions of their doctors' notes and after-visit summaries: a mixed methods study of patients at safety-net clinics. Health Expect 2018;21(2):485-493 [FREE Full text] [doi: 10.1111/hex.12641] [Medline: 29095554]
25. Lyles CR, Gupta R, Tieu L, Fernandez A. After-visit summaries in primary care: mixed methods results from a literature review and stakeholder interviews. Fam Pract 2019;36(2):206-213. [doi: 10.1093/fampra/cmy045] [Medline: 29846584]
26. Swedlund M, Kamnetz S, Birstler J, Trowbridge E, Arndt B, Micek M, et al. Reduction in medication refill encounters through primary care redesign workflow changes. J Ambul Care Manage 2022;45(1):36-41. [doi: 10.1097/JAC.0000000000000398] [Medline: 34690304]
27. Vordenberg SE, Zikmund-Fisher BJ. Older adults' strategies for obtaining medication refills in hypothetical scenarios in the face of COVID-19 risk. J Am Pharm Assoc 2020;60(6):915-922.e4 [FREE Full text] [doi: 10.1016/j.japh.2020.06.016] [Medline: 32680780]
28. Moreno G, Lin EH, Chang E, Johnson RL, Berthoud H, Solomon CC, et al. Disparities in the use of internet and telephone medication refills among linguistically diverse patients. J Gen Intern Med 2016;31(3):282-288 [FREE Full text] [doi: 10.1007/s11606-015-3500-6] [Medline: 26311200]
29. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
30. Laukka E, Huhtakangas M, Heponiemi T, Kujala S, Kaihlanen A, Gluschkoff K, et al. Health care professionals' experiences of patient-professional communication over patient portals: systematic review of qualitative studies. J Med Internet Res 2020;22(12):e21623 [FREE Full text] [doi: 10.2196/21623] [Medline: 33289674]
31. Grossman LV, Masterson Creber RM, Benda N, Wright D, Vawdrey D, Ancker J. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
32. Medication safety tips for seniors. Oklahoma State Department of Health. 2020. URL: https://oklahoma.gov/content/dam/ok/en/health/health2/aem-documents/prevention-and-preparedness/injury-prevention/falls-prevention/up-medication-safety-fact-sheet-2020.pdf [accessed 2023-05-30]
33. Defanti e Souza F, da Silva Santana C. A descriptive study about the use of pillboxes by older adults. Health 2013;05(12):103-109.
34. Mehdinia A, Loripoor M, Dehghan M, Heidari S. The effect of pillbox use on medication adherence among elderly patients: a randomized controlled trial. Int Electron J Med 2020;9(1):38-43. [doi: 10.34172/iejm.2020.07]
35. Cross AJ, Elliott RA, Petrie K, Kuruvilla L, George J. Interventions for improving medication-taking ability and adherence in older adults prescribed multiple medications. Cochrane Database Syst Rev 2020;5(5):CD012419 [FREE Full text] [doi: 10.1002/14651858.CD012419.pub2] [Medline: 32383493]
36. National institute on aging. Taking Medicines Safely as You Age. URL: https://www.nia.nih.gov/health/taking-medicines-safely-you-age [accessed 2023-05-30]
37. Svensk J, McIntyre SE. Using QR code technology to reduce self-administered medication errors. J Pharm Pract 2021;34(4):587-591. [doi: 10.1177/0897190019885245] [Medline: 31835959]
38. Carvalho TPD, Okuno MFP, Campanharo CRV, Lopes MCBT, Batista REA. Patients' knowledge about medication prescription in the emergency service. Rev Bras Enferm 2018;71(2):329-335 [FREE Full text] [doi: 10.1590/0034-7167-2017-0002] [Medline: 29412290]
39. Kim JM, Suarez-Cuervo C, Berger Z, Lee J, Gayleard J, Rosenberg C, et al. Evaluation of patient and family engagement strategies to improve medication safety. Patient 2018;11(2):193-206. [doi: 10.1007/s40271-017-0270-8] [Medline: 28795338]
40. Lee M, You M. Effects of COVID-19 emergency alert text messages on practicing preventive behaviors: cross-sectional web-based survey in South Korea. J Med Internet Res 2021;23(2):e24165 [FREE Full text] [doi: 10.2196/24165] [Medline: 33544691]
41. Kim AM, Bae J, Kang S, Kim Y, Lee J. Patient factors that affect trust in physicians: a cross-sectional study. BMC Fam Pract 2018;19(1):187 [FREE Full text] [doi: 10.1186/s12875-018-0875-6] [Medline: 30497400]
42. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
43. Donohue K, Katok E, Leider S. The Handbook of Behavioral Operations. New Jersey, United States: John Wiley & Sons, Inc; 2018.
44. Lang Y, Chen KY, Wu Y. Learning under the inventory problem of economic order quantity: a behavioural study. Int J Inventory Res 2023;1(1):1. [doi: 10.1504/ijir.2023.10054788]
45. Walters K, Christakis DA, Wright DR. Are mechanical turk worker samples representative of health status and health behaviors in the U.S.? PLoS One 2018;13(6):e0198835 [FREE Full text] [doi: 10.1371/journal.pone.0198835] [Medline: 29879207]

## Abbreviations

**AHRQ:** Agency for Healthcare Research and Quality
**IRB:** institutional review board
**MTurk:** Amazon Mechanical Turk
**PROMIS:** Partnership for Resilience in Medication Safety
**SoPHIE:** Software Platform for Human Interaction Experiments

*Edited by T de Azevedo Cardoso; submitted 20.03.24; peer-reviewed by J Su, G Lisa, R Marshall; comments to author 15.04.24; revised version received 21.05.24; accepted 10.07.24; published 12.08.24.*

*Please cite as:*
*Lang Y, Chen KY, Zhou Y, Kosmari L, Daniel K, Gurses A, Young R, Arbaje A, Xiao Y*
*Perception of Medication Safety–Related Behaviors Among Different Age Groups: Web-Based Cross-Sectional Study*
*Interact J Med Res 2024;13:e58635*
*URL: https://www.i-jmr.org/2024/1/e58635*
*doi:10.2196/58635*
*PMID:*

Original Paper

# Medication Management Strategies to Support Medication Adherence: Interview Study With Older Adults

Lisa Gualtieri[1], PhD, ScM; Mathilda Rigby[2], BS; Deelia Wang[1]; Elaine Mann[1], MD

[1]Tufts University School of Medicine, Boston, MA, United States

[2]Harvard Medical School, Boston, MA, United States

**Corresponding Author:**
Lisa Gualtieri, PhD, ScM
Tufts University School of Medicine
136 Harrison Ave
Boston, MA, 02111
United States
Phone: 1 7813309456
Email: lisa.gualtieri@tufts.edu

## Abstract

**Background:** Home medication management has been insufficiently studied, including the factors that impact the development and effectiveness of adherence strategies under both routine and anomalous circumstances. Older adults are a particularly important population to study due to the greater likelihood of taking medication in combination with the desire to "age in place."

**Objective:** This interview study aims to understand how older adults develop medication management strategies, identify when and why such strategies succeed or fail, learn more about how older adults think about their medication, and explore interventions that increase medication adherence.

**Methods:** This study used a qualitative, semistructured interview design to elicit older adults' experiences with home medication management. Overall, 22 participants aged ≥50 years taking 1 to 3 prescription medications were recruited and interviewed. Interview responses were recorded, and thematic, qualitative analysis was performed by reviewing recordings and identifying recurring patterns and themes. Responses were systematically coded, which not only facilitated the identification of these themes but also allowed us to quantify the prevalence of behaviors and perceptions, providing a robust understanding of medication management and medication adherence.

**Results:** Participants reported developing home medication management strategies on their own, with none of the participants receiving guidance from health care providers and 59% (13/22) of the participants using trial and error. The strategies developed by study participants were all unique and generally encompassed prescription medication and vitamins or supplements, with no demarcation between what was prescribed or recommended by a physician and what they selected independently. Participants thought about their medications by their chemical name (10/22, 45%), by the appearance of the pill (8/22, 36%), by the medication's purpose (2/22, 9%), or by the medication's generic name (2/22, 9%). Pill cases (17/22, 77%) were more popular than prescription bottles (5/22, 23%) for storage of daily medication. Most participants (19/22, 86%) stored their pill cases or prescription bottles in visible locations in the home, and those using pill cases varied in their refill routines. Participants used ≥2 routines or objects as triggers to take their medication. Nonadherence was associated with a disruption to their routine. Finally, only 14% (3/22) of the participants used a time-based reminder or alarm, and none of the participants used a medication adherence device or app.

**Conclusions:** Participants in our study varied considerably in their home medication management strategies and developed unique routines to remember to take their medication as well as to refill their pill cases. To reduce trial and error in establishing a strategy, there are opportunities for physicians and pharmacists to provide adherence guidance to older adults. To minimize the impact of disruptions on adherence, there are opportunities to develop more durable strategies and to design aids to medication adherence that leverage established daily routines.

***(Interact J Med Res 2024;13:e53513)*** doi:10.2196/53513

**KEYWORDS**

home medication management; medication adherence; prescription drugs; adherence devices; adherence apps; pill cases; aging in place; independent living; aging; medication; older adults; prescription; interview; interview design; design; app; mobile phone

# Introduction

## Background

Medication adherence—defined as taking medication as agreed with one's prescriber—is a crucial part of aging well. Studies have shown that ≥50% of US adults do not take their prescriptions as directed and that medication nonadherence is responsible for as many as 33% to 69% of hospital admissions and 125,000 deaths annually [1-3]. The World Health Organization emphasizes the importance of medication adherence, stating that "[i]ncreasing the effectiveness of adherence interventions may have a far greater impact on the health of the population than any improvement in specific medical treatment" [4].

There is a rising rate of medication nonadherence across all ages, sex, and race groups in the United States [5]. The greatest concern, however, is for older adults. According to population estimates by the United Nations World Population Prospects report, 1 in every 4 persons will be aged ≥65 years by 2050 in Europe and Northern America [6]. This represents a doubling in the older adult population—approximately 9% in 2019 to an estimated 16% in 2050. Aging populations require consistent medication adherence for health maintenance; approximately 67% of US adults aged 45 to 64 years take at least 1 prescription drug, and this figure rises to 88.5% for those aged ≥65 years [7].

Older adults are underrepresented in research on medication adherence and may face some unique barriers compared to the average population [8,9]. There is a consistently strong preference to "age in place" by older populations, preferring to remain within communities instead of institutional care [10]. However, home settings are associated with medication nonadherence [8,11]. Identifying barriers to medication adherence related to home medication management, specifically related to unintentional nonadherence, may lead to successful strategies to aid older adults who wish to age in place [12]. Forgetting to take medication is one of the most common contributors to unintentional medication nonadherence and is particularly relevant for older adults desiring to live independently [12].

Many interventions have been designed to aid in individual medication adherence, generally specific to a disease or treatment. Innovation in general prescription packaging, distribution, and education has been limited. Some innovations in container, labeling, and package design include blister packs, PillPack, and a variety of adherence devices ranging from digital pill containers to dispensing devices [13-16]. Studies have shown limitations to many of these innovations, particularly in their design and usability. These innovations have also had limited commercial success. Overall, no substantial impact on adherence has resulted from the use of simple adherence devices [17,18]. More than 700 medication adherence apps exist, but users often report technical difficulties such as schedule inflexibility because many do not take, or need to take, their medication at an exact time, as well as notification fatigue from daily timed reminders to take their medication [19,20]. Therefore, innovation in containers, packaging, devices, and apps represents potential targets for improvement of adherence [21-24].

Another way to improve medication adherence is counseling patients on home medication management practices [25]. Health care providers can play an efficacious role in encouraging behaviors to support adherence, especially ones tailored to patients' medication regimens and needs [20,26]. External influences such as reduced time for medical appointments and the increased use of mail delivery of prescriptions, necessitating fewer pharmacy visits, reduce opportunities for medication adherence guidance from physicians or pharmacists, respectively.

Medication adherence often relies on the development of a behavior that is repeated in response to triggers and fits into a routine [27]. Triggers are actions that are taken or objects that are encountered that can help patients remember to take their medication. Habit formation is a main determinant of behavior change but can take trial and error to be established, especially when no guidance is provided by a health care provider or suggested from past caregiving experience [28]. Selection of triggers and formation of habits specific to medication management are understudied, especially regarding the factors that may impact the development and sustainability of strategies within the context of both routine and anomalous circumstances.

## Objectives

This study aimed to identify home medication management practices used by community-dwelling older adults with simple medication regimens. Specifically, this qualitative study aimed to (1) understand how older adults develop medication management strategies, (2) identify strategies that lead to adherence and when and why such strategies succeed or fail, (3) learn more about how older adults think about their medication, and (4) explore interventions that increase medication adherence.

# Methods

## Approach

On the basis of the results of a survey [29], we planned an interview study to learn in depth about the medication management experiences of older adults in their homes. Our goal was to better understand how older adults fit medication taking into their daily lives. Hence, our interest was in patients with simple regimens involving 1 to 3 medications because, with more, medication management is a more all-consuming daily task.

We developed an interview guide and semistructured interview process to elicit responses from older adults on their experiences with home medication management, including selection of medication storage location and development of medication-taking routines, to characterize which factors positively influence medication adherence when managing a relatively small number of medications. Many interview questions were based on survey results, which showed what respondents did but did not reveal why or how they made their decisions [29]. Therefore, the interview questions were designed to elicit a richer narrative of how and why home medication

management decisions are made and patients' perceptions about what worked and what did not. The protocol was piloted on Zoom (Zoom Video Communications, Qumu Corporation) with 1 individual in the study demographic to test timing, flow, and wording, and no changes were made. The interview guide is included in Multimedia Appendix 1.

The interview questions were developed to elicit information about how older adults manage their medications in their homes as part of the daily routine and under anomalous circumstances such as travel. Interview questions asked participants to recount their medication routines and were designed to cover the 3 components of adherence—initiation, implementation, and discontinuation—to fully address the stages of adherence [30].

For the initiation phase, participants were asked who prescribed their medications, whether this was during in-person or virtual visits, and if they visited a pharmacy or used mail delivery. Participants were questioned about the extent of their health care provider's counseling or advice on home medication management, including how and where to store prescriptions and how to remember to take them. For the implementation phase, patients were asked about their daily home medication management practices, including how they administer, store, and identify their prescription medications. They were questioned about any vitamins or supplements they take. They were asked about any stigma they experienced about taking medication or about visitors to their home seeing their medication. In addition, participants were questioned about their strategies for ascertaining adherence, including alarms, apps, devices, and assistance from others, and their interest in the use of a medication adherence device that used sensors to alert them only when they forgot to take their medication. This device was described as using the same principle as a seat belt, chiming only when you both forgot to latch it and turn on the ignition in a car. The device was described as using one sensor on the participant's prescription bottle or pill case and another sensor on a coffee pot, toothbrush, or another object related to the participant's daily routine. Finally, in the discontinuation phase, participants were questioned about nonadherence. Demographic questions were asked at the end.

## Participants

Participants included older adults recruited in July 2022 through the Osher Lifelong Learning Institute (OLLI) at Tufts University [31]. Participants were recruited through postings in the weekly email newsletter sent to approximately 2000 OLLI members who subscribed. Inclusion criteria were participants aged ≥50 years, taking 1 to 3 prescription medications for chronic conditions (as opposed to taking on a short-term or an as-needed basis), proficient in English, and with no cognitive impairment.

A total of 22 people responded to the recruitment materials using a link in the OLLI newsletter. All the 22 respondents consented to the web form, responded to the email to schedule a time for an interview, and participated in the study. Ages ranged from 56 to 87 (mean 70.5, SD 6.3) years, and 82% (18/22) of the participants identified as female and 18% (4/22) as male. All participants were White, and all had a bachelor's degree or higher educational status. All were community dwelling, living in a house or apartment alone or with a partner, and all lived in an urban or suburban setting in Greater Boston.

Despite the constraints imposed by a small sample size, data saturation was largely achieved. This was evident by recurring patterns and themes emerging in the responses and the ability to draw meaningful conclusions from our sample.

Semistructured qualitative interviews were conducted with 22 participants during August 2022. The study was conducted during the COVID-19 pandemic, and owing to physical distancing protocols, all interviews were conducted via Zoom teleconferencing software. The duration of each interview was 30 to 45 minutes. Consent to participate was obtained at recruitment, and consent to record the interview was obtained at the start of each interview.

## Interview Team

The interviews were conducted by 1 researcher (LG), who had many years of prior experience conducting interviews. Emails were sent to arrange an interview time, and at the start of each interview, participants were told that the purpose of the interview was for a research project on medication adherence and were asked to consent to recording the session. No bias, assumptions, or interests in the topic were reported about the interviewer. Another researcher (MR) attended all sessions but did not conduct any interviews. Both researchers (LG and MR) took informal notes during the sessions. Recorded sessions were transcribed verbatim.

## Ethical Considerations

Study protocols were approved by the Tufts University Health Sciences Institutional Review Board (STUDY00002865). All participants consented to participate in the interviews and were told the purpose of the research in the consent form. All participants were asked at the start of the interview if they would consent to recording the interviews, that they would be only identified by a participant number and their names or any identifying information would not be used, that all recordings would be securely stored and erased at the culmination of the research, and that they could skip any question they did not wish to answer. The recordings and all data were identified only by a participant number and were thus deidentified. All participants were compensated with a US $50 Amazon gift card, which was sent to their email address as a way of showing our appreciation for their time. This amount was deemed appropriate for 30 to 45 minutes without being coercive.

## Analysis

The transcript analysis aimed to explore and understand the experiences of the participants; therefore, thematic analysis was chosen as the analytic strategy. Thematic analysis is a qualitative descriptive approach that is used to identify, analyze, and report patterns within data and is useful for analyzing narratives [32]. Thematic analysis was performed by 2 researchers reviewing and coding recordings and transcripts. A subset of interview questions, deemed most relevant to uncovering home medication management practices, were coded as part of this analysis.

To initiate the analysis, the researchers reviewed the interview responses to identify prominent themes and noteworthy topics

emerging from the data. After identifying potential areas of interest, a structured list of questions and topics was developed to serve as a preliminary coding framework for the analysis. The researchers independently reviewed the recordings and interview notes to apply the initial coding framework. They engaged in regular meetings to compare their findings and discuss any discrepancies. This collaborative approach ensured consistency and reliability in theme identification. It also allowed for the iterative refinement of the coding scheme through both deductive and inductive methods, as themes were allowed to evolve naturally from the data, while the initial framework provided a guide to maintain focus on the study's aims.

Coding the responses facilitated the identification of significant themes and allowed the research team to quantify the prevalence of behaviors and perceptions, providing a robust understanding of medication management and medication adherence.

# Results

## Sample Description

Five themes were identified relevant to home medication management strategies. These themes corresponded to the stages of medication adherence: initiation, implementation, and discontinuation.

### *Theme 1: Participants' Experiences of Obtaining Medications*

The first set of interview questions asked participants how their medications were prescribed and obtained. The purpose of these questions was to start the interview with straightforward questions that were simple to answer yet relevant to this research. An additional objective was to ask participants if they received counseling or advice from a physician or pharmacist regarding home management of medication, including where to store them or how to remember to take them.

In response to being asked about who prescribed their medications, how they obtained them, and if any counseling or advice was provided, all the participants obtained prescriptions from their primary care physician or specialist, yet none received guidance from a health care provider about how to devise an effective medication management strategy. More than half of the participants (13/22, 59%) used trial and error to develop a strategy, which included trying different locations in the home or trying a pill case after seeing one in a local pharmacy. Five participants devised a strategy based on experience assisting someone else manage their medication. One participant who devised a strategy based on prior experience stated the following:

> *I have managed medications in the past for my mother and aunt, both of whom are deceased, but they had pill cases [which are] pretty common. It seems [like] it's a good organizational tool. I didn't think a whole lot about it [and] went out and got a pill case when I first started having prescriptions.* [Participant 13]

The remaining 4 participants devised their medication management strategy using suggestions from a friend or family member.

Approximately half of the participants (10/22, 45%) received prescriptions by mail delivery only, 14% (3/22) of the participants used a combination of mail and pharmacy pickup, and the remaining 41% (9/22) used pharmacy pickup only. All but 1 participant (21/22, 95%) received a 90-day supply of their prescription medication.

### *Theme 2: Participants' Experiences of Taking Medications at Home*

All participants (22/22, 100%) responded to the question, "Can you walk us through your daily schedule for taking your medications, specifically when you take your medications and where you store them?" All participants provided a description of their home management practices, which constituted unique routines. All participants reported on what they used for medication storage and where containers were placed, refill strategies for pill cases, and how they remembered to take their medication.

For medication storage, only 23% (5/22) of the participants kept their medication in the prescription bottles it was received in. Most of the participants (17/22, 77%) used pill cases to store their medication. Of these, there was considerable variance in the type of pill case used, including the number of compartments. Of the 17 participants, 7 (41%) used 1 weekly pill case with 7 compartments. For participants with morning and evening medication, 6% (1/17) used a 14-compartment pill case, while 29% (5/17) used 2 separate pill cases each week. Two participants 12% (2/17) used 2 pill cases each to be able to refill them together for a 2-week medication supply, 1 of them citing the inconvenience of refilling pill cases and the desire to do so as infrequently as possible. Referring to their pill case, one participant describes the following:

> *It's kind of a pain to fill so I kind of put it off...but if it's obviously empty [or if] there's maybe one slot to go, I'll say awesome, let's just fill it...I've got some time.* [Participant 5]

One participant 6% (1/17) had a separate pill case for each of their medications, and another participant 6% (1/17) used a pill case that stored a 4-day supply of medication.

Most US pill cases are designed with "Sunday" on the left-hand side; hence, we expected participants to have a weekend refill routine; instead, there was variance in when pill cases were refilled and how participants remembered to refill them. Overall, 41% (7/17) of the participants relied on the visual cue of an empty pill case to refill it; however, this method did not typically lead to them refilling their case on the same day every week. One participant stated the following:

> *When I get to the point where something is empty, and I say, oh, time to refill. It should be every seven days, but sometimes I might forget. Or I don't know why. But it doesn't always work out to be every seven days. But anyway, whenever they're empty, then they need to be refilled.* [Participant 7]

Furthermore, 59% (10/17) of the participants consistently refilled on a specific day of the week. One described the routine for refilling as follows:

> *Sunday morning, it's a routine. After breakfast, I drag [my medication] out. I have two different sets [of pill cases]. So I always have one in reserve in the closet with the bottles of the pills. I can easily take [the medication] on Sunday. I don't have to wait till I fill them in order to take pills [since] I already have a set ready. Usually after breakfast or when I have a chance during the day on Sunday, I'll go ahead and fill the one that I've just emptied in the previous week and put it in the closet.* [Participant 17]

Participants using prescription bottles used 1 storage location for their currently used bottle; those who received a 90-day supply in multiple bottles used a secondary location for excess. Participants using pill cases used 2 storage locations: a primary storage location for the pill case itself and a secondary storage location for the prescription bottles used to refill the pill case. The primary storage locations used for prescription bottles and pill cases were visible locations in their home for 86% (19/22) of the participants. The kitchen table was the most common primary storage location (10/22, 45%), and the bathroom counter was the second most common primary storage location (4/22, 18%). For the 17 participants using pill cases, 14 (82%) used a secondary storage location that was hidden from sight. The most common secondary storage location was the kitchen cabinet (7/22, 32%), followed by the bathroom cabinet (5/22, 23%).

Independent of the storage container used, most of the study participants (20/22, 91%) took their medication during a time range tied to a routine, such as eating a morning meal, while only a small number (2/22, 9%) reported taking medication at an exact time every day. One participant reported taking her evening medication as follows:

> *Sometime after dinner and before going to sleep. It's probably a two-, three-, or four-hour range there.* [Participant 11]

Only a few participants (3/22, 14%) used a digital time-based reminder or alarm to manage their medication. One participant set an Alexa device to give her an oral reminder to take her pill at 10 AM every day, 1 participant used a smartphone reminder, and 1 participant used an alarm. None of the participants used medication adherence devices or apps for reminders.

All participants (22/22, 100%) relied on at least 2 triggers to remind them to take their medication. No 2 medication-taking routines were identical among participants. Action triggers included eating a meal (10/22, 45%), getting ready for bed (5/22, 23%), and brushing teeth (4/22, 18%); object triggers included a visible pill case (17/22, 77%) and a water glass (4/22, 18%). Because most participants stored their medication in a visible area, they could see their medication container as they engaged in a routine; for example, of the 10 participants who took their medication during a meal, 8 (80%) stored their prescription bottle or pill case on a kitchen table or counter. Storing their medication in a visible area served as a second trigger, backing up the routine-based trigger. One participant recounted the following:

> *I have what I call my staging area, which is an area between my kitchen and my dining room. [My medication] stays in the [staging area] and since I take that medication right after dinner it's right there. As I'm clearing the table, after I put the dishes in the sink, I just go and I take the medication right after dinner, and it's visibly right there.* [Participant 6]

Some triggers were tactile, not just visible. One participant used the spatial orientation of her pill bottles to manage medication adherence, reporting the following:

> *I came up with a scheme, where I keep the medicines on one side of my microwave, or my toaster oven. When I take it, I put it on the other side.* [Participant 4]

While all participants relied on at least 2 triggers to remind them to take their medication, more than half (15/22, 68%) of the participants relied on ≥3 triggers. One participant who relied on 3 triggers, taking medication with a meal, using a pill case, and placing it on the dining table, missed the first trigger but saw their pill case, which acted as a fallback reminder.

### *Theme 3: Factors Contributing to Nonadherence*

From the participants' descriptions of their medication management, the use of ≥2 triggers served as "a safety net" most of the time, providing multiple reminders to take medication. Yet even multiple triggers failed at times. The most reported reason for medication nonadherence among participants was an unplanned or unexpected change of routine (13/22, 59%), such as missing breakfast, waking up later in the day, or being distracted by a phone call. Such events typically led to the absence of, or overlooking, a specific trigger. For example, a participant who relied on eating breakfast as a cue reported forgetting to take her medication if she skipped breakfast:

> *If I have to go somewhere, first thing in the morning, that's a typical time when I forget. Because sometimes I don't even have time for breakfast or for one reason or another didn't get around to it. Then the next day, it's Monday, but I'm looking at the Sunday [compartment of the pill] case saying, "Oh, I guess I forgot to take it yesterday."* [Participant 7]

Another participant acknowledged difficulties when her usual routine was interrupted or altered, reporting the following:

> *I just got distracted. I was on the phone with a friend. [Forgetting medication is] more apt to happen if I'm with my mom or somewhere other than in my own home because I'm out of the routine, even though I have the pill with me.* [Participant 17]

The second most reported reason for nonadherence was travel. The 8 participants who reported nonadherence during travel explained it by citing a change in schedule, not being able to store their medication in the same place, or not having the same triggers in their usual routine available. One participant recalled the following:

> *I have occasionally forgotten. Frankly, it's when I'm on vacation; even though I have them in the [weekly pill case], my routine is different on vacation. It [the*

> *pill case is] not in my kitchen on vacation. I'm away someplace.* [Participant 19]

Another participant reported forgetting to pack his medication:

> *There was one time, I remember, when I left my pills home. So there were like three days where I was not taking the pills.* [Participant 18]

Stigma was not reported as a contributor to nonadherence. No participants reported experiencing stigma about taking medication or about visitors to their home seeing medication. When asked about stigma, respondents spoke about how common it was for older adults they knew to be on medication, which, to them, eliminated the experience of stigma or embarrassment.

### *Theme 4: Perception of Medication*

Participants were asked about how they thought about their medication, for example, prescription name, pill purpose, or pill appearance. In response to this question, approximately half of the participants (10/22, 45%) thought about their medication by the chemical name, with the next most popular response being by the appearance of the pill (8/22, 36%). A smaller number (2/22, 9%) thought about the medication's purpose or the medication's generic name. One participant described medications as follows:

> *I can't name the one that I have been taking the longest, which is upstairs in the bathroom. [The medication for] underactive thyroid I just remember by the name Levoxyl. And I've always remembered it. The one downstairs I just recently started and it's a statin. I still don't know what the name of it is. But it's just for cholesterol.* [Participant 14]

Another participant described medications as follows:

> *I know all of them by their generic names. And when they're new, I think about what they're meant to do. But over time, I recognize them by their shape and color. On a day to day basis, I probably look for the shapes. And it really is off-putting when the pharmacy either changes the generic or I change my insurance plan and deal with the pharmacy benefits manager who happens to have a different generic.* [Participant 20]

Furthermore, of the 19 participants who took vitamins or supplements in addition to their prescription medication, 18 (95%) treated their medications the same, not distinguishing between what was prescribed by a physician, recommended by a physician, or something they were taking independent of their physician. These 18 participants stored their vitamins and supplements in the same pill case as prescription medication, and they integrated them into their medication routine, rather than distinguishing between their prescription medications and their vitamins and supplements.

### *Theme 5: Interest in Adherence Device*

The final question to participants was about their interest in the use of a medication adherence device that used sensors to alert them only when they forgot to take their medication. Of the 22 participants, 17 (77%) were interested in using this device; 12 (71%) participants expressed interest in using the device immediately, whereas 5 (29%) were open to using the device if their medication routine got more complicated or they experienced any cognitive decline. Of those 5 participants, 1 (20%) talked about the need to use a device in the future, saying the following:

> *At this point now, I wouldn't [be interested in a device], because I just don't need it. But certainly, if I was struggling to remember to take them or if I had, like some people I know, this very, very complex regimen. So, I would be open to it at some point, but not now.* [Participant 9]

Out of the 5 participants who were not interested in the device, 3 (60%) explained that they saw no need to change or add to their current routine, while 2 (40%) rejected the idea due to the potential of notification fatigue or an unwillingness to use technology for medication adherence.

## Discussion

### Principal Findings

The purpose of this qualitative study was to explore the experiences of older adults managing their medication in their homes. In conducting this study, our most significant finding was the complexity and uniqueness of what participants did to manage simple medication regimens. Another significant finding was the extent to which trial and error or prior experience were used to develop strategies without guidance from health care professionals.

Medication management strategies need to encompass how people think about their medication and not the artificial demarcation of prescription medication only. Study participants who took vitamins and supplements thought about their medication as prescriptions, vitamins or supplements recommended by physicians, and vitamins or supplements recommended by a friend or another source.

Another finding was the variability in the use of pill cases, including the number and type of pill cases used and the frequency and timing to refill them. Related to this was the complexity of medication storage location selection for both primary storage, for example, a storage location accessed daily for medication, and secondary storage, for example, storage used for extra medication supply. Overall, the sheer variability in home medication management strategies was surprising and unexpected, especially for participants with relatively simple medication regimens.

Additional results from each theme are discussed in the subsequent sections.

### Theme 1: Participants' Experiences of Obtaining Medications

All study participants expressed that they received no guidance from a physician or pharmacist on any aspect of home medication management, including how to establish a routine to be adherent. Of the 22 participants, 13 (59%) designed their own medication management regimens. A small number relied on prior experience helping someone else manage their

medication or through the advice of a friend or family member. This lack of guidance from health care providers presents a missed opportunity to increase medication adherence from a trusted professional [30]. A survey conducted by Gualtieri et al [29] found that 96% of middle-aged and older adult respondents were receptive to receiving guidance from a physician or pharmacist regarding their medication management. This guidance could occur as part of prescribing or a medication review by a physician or by a pharmacist during prescription pickup.

With more than half of participants receiving prescriptions by mail only (10/22, 45%) or a combination of mail and pharmacy pickup (3/22, 14%) and approximately all receiving a 90-day supply of their prescription medication, fewer pharmacy visits are required as part of obtaining prescriptions. While this may decrease the likelihood of running out of medication, it also serves to decrease the need to enter a pharmacy location and therefore may reduce the extent to which an older adult establishes a relationship with or asks questions of a pharmacist. Less time in a pharmacy may limit opportunities to explore adherence-related tools, such as pill cases, which are often displayed next to or near pharmacies in retail stores. Less explored consequences of the rise in mail delivery may be theft or degradation of medicine by extreme temperatures, rain, or humidity.

### Theme 2: Participants' Experiences of Taking Medications at Home

Participants described a wide range of routines for taking their medication. Each routine was unique with varying degrees of complexity. Most integrated multiple tactile and visual triggers into their daily routines to prompt them to take their medication each day. These adherence regimens may be so varied due to a lack of physician or pharmacist guidance or established norms for home medication management. Additional diversity was found in the unique locations where participants stored their medication and the timing with which they refilled it.

When participants described their experiences of home medication management, they included how they stored their pills. Pill cases were the most popular medication adherence device used at home. A possible reason for the high prevalence of weekly pill cases is that they provide direct feedback on whether someone took their medication using visual cues. Whether the slot in the pill case is empty or full is a straightforward indicator, unlike a prescription bottle, where users have no way of knowing if they took a pill that day unless they count pills. Weekly pill cases are also the most commonly seen medication adherence tool at pharmacies, though as evidenced by the interviews, the style and how they are used can vary greatly. This theory aligns with other literature, which demonstrates that pill boxes can effectively aid medication adherence. Another study on medication adherence found that some patients favored using pill boxes for managing their medications because they provided a visual reminder that they had taken their doses [33].

The most common primary storage location for medication was the kitchen, which was associated with using food preparation or consumption as a trigger for medication taking. This may be because mealtimes are a stable component of a daily routine. Another possible factor explaining the popularity of a kitchen storage location is that the participants enter the kitchen to complete specific tasks, as opposed to using it for long periods. This prescription bottle or pill case may act as a visual cue in the kitchen.

### Theme 3: Factors Contributing to Nonadherence

Participants, in describing their medication management strategies, referred to the use of multiple triggers that served as reminders to take their medication. Because triggers sometimes fail, they may depend on their specific context. An object or action without its usual context might be less effective as a reminder.

When asked to recount the last time they were nonadherent, participants described a disruption leading to a change in routine in their home or being in a different location due to travel. Because the recounted occurrences are events that are not under the control of the participant, developing a robust medication management strategy should ideally accommodate these anomalies. During a change in routine, a trigger that is usually relied upon may be absent. More robust routines with multiple triggers may endure disruptions better, or more durable triggers that are not disrupted by unexpected events are needed. More planning may also be needed to accommodate disruptions and unplanned events. Adherence devices may provide reminders to reduce nonadherence under anomalous circumstances.

Our study pinpointed factors contributing to nonadherence that are consistent with those identified in an interview study by Mickelson and Holden [34]. Their findings highlighted that disruptions in daily routines or travel could result in lapses in medication adherence. In addition, they observed a variety of tools and technologies being used to manage medications, including pill boxes. Uniquely, their study also revealed that stigma played a role in nonadherence; participants were hesitant to be seen as sick, leading them to skip medications in social settings, a finding that contrasts with our study [34]. Our participants noted that medication use among older adults is so common that it does not carry any stigma for them.

### Theme 4: Perception of Medication

How participants referred to their medication varied considerably, with most referring to their medication using the chemical name, followed by appearance. The use of chemical names could be due to education because all participants had obtained a bachelor's degree or higher. Another factor influencing how people refer to their medication could be the number of medications that they take; it may be harder to keep track of chemical names as the number of medications increases, especially with the complexity of chemical names.

The second most common mode of reference participants used was appearance. When pills are similar in appearance, concerns about medical error arise. It is worth investigating the correlation between how people think of their medication and medication adherence and if discussion with a physician or pharmacist should emphasize the medication name, appearance, and purpose to aid in accurate identification by patients, especially for tasks such as filling a pill case. The primary concern for medical error

may be when a patient who relies on appearance for prescription identification receives a new generic or a dosing change and the new pill is a different shape, size, or color.

Most participants thought of their medication as prescription medication, vitamins, and supplements, without categorizing differently what they were prescribed, recommended by a physician, or took of their own volition. Adherence is more critical for prescription medications, but strategies to guide patients should ideally address how they think about their medication regimen and, for pill case users, how they select, use, and refill pill cases.

### Theme 5: Interest in Adherence Device

Only 1 participant currently relied on a general, as opposed to adherence, technology to assist with medication; this participant relied on a daily reminder from Alexa. None of the study participants currently used adherence devices or apps to generate reminders. More than two-thirds of the participants (17/22, 77%) expressed interest in using a device to assist with adherence only when needed, in contrast to devices with timed reminders. Only 2 individuals expressed a lack of interest in this technology. In addition, most study participants lived alone without others to remind them to take their medication, providing an opportunity where an adherence device might be beneficial as a backup. A lack of experience, positive or negative, with medication adherence device or app use did not seemingly deter participants from expressing interest in a device that integrated naturally into their routine.

## Strengths and Limitations

The strength of participant recruitment through the OLLI at Tufts University was that our recruitment goal of ≥20 participants was met quickly. However, the disadvantage was that our sample was not representative of the US Census for adults aged >50 years. The narrowness of our study population, due to our recruitment strategy, is a limitation of this study. A subsequent interview study was conducted with adults who identify as racial or ethnic minority individuals.

While our cohort consisted of participants who were more educated, and hence likely to have higher health literacy skills, they still experienced unintentional nonadherence. A higher level of education is associated with higher socioeconomic status; however, none of the strategies deployed for medication adherence by participants were costly. Furthermore, many participants were not working full time or were retired, which may influence their daily routines and their ability to travel. Because most of the participants lived alone, they were less likely to be able to rely on others to remind them.

Finally, participants were limited to those taking 1 to 3 medications and experiencing no cognitive decline. A more representative sample of older adults should include those taking more medications, including those with complicated schedules. It should also include those who are experiencing minor cognitive decline yet are living independently because this might add additional barriers to medication adherence strategy development and may increase forgetfulness.

Despite some limitations in the diversity of participants, this study has strengths in uncovering home medication strategy development and execution through the set of questions eliciting participants' experiences. Face-to-face interviews lasting 30 to 45 minutes through Zoom enabled us to gain more insight and details than we would have discovered through a survey; in fact, our prior survey work led to the design of an interview guide to gain a deeper understanding. Participants shared freely how they developed and executed a home medication management strategy, when the strategy was successful, and when and why it failed.

## Future Directions

Future work will build on the results of this study as well as the completed interview studies on racial and ethnic minorities. One goal is to create educational interventions that lead to improved adherence under routine and anomalous circumstances and eliminate or reduce the trial and error process so often used. Ideally, an educational intervention would be delivered by physicians when prescribing, by pharmacists when dispensing, or by either professional when nonadherence is indicated by a patient or by refill frequency.

Related to this, and in support of patient-physician communication about adherence, another goal is to develop a scale for home medication management, much like those for medication adherence but focused on home practices. This scale could be used by physicians when prescribing medication to determine how likely a patient is to be adherent based on their current lifestyle or by pharmacists or family caregivers to provide guidance in developing an individualized medication management strategy.

Another goal is to design and test medication adherence devices that take a failsafe approach to helping older adults live independently longer. While current medication adherence devices and apps rely on time-based triggers, our study shows the reliance on routine, leading to our interest in routine-based triggers. Our study found that participants managed their medication in the context of a routine, yet their routines did not take place at an exact time but rather upon rising or in the morning. Alerts or reminders set for a specific time may be disruptive if the timing of the routine is variable. They can be even more confusing if the person has already taken their medication but still receives a reminder to do so. This may result in notification fatigue or in turning off notifications altogether. Thus, we plan to design and test a medication adherence device that reminds users to take their medication only when their current routine fails them and is robust enough to accommodate disruptions. Given the high interest from our participants when asked about this new medication adherence device, this has the potential to reduce unintentional nonadherence in older adults who are aging in place.

Given that study participants' medication management strategies treated prescribed medication the same way as vitamins and supplements, we will further explore the implications of this for adherence. Home medication management strategies, to be effective, should align with how patients think about their medication. This should include any medication, prescription, or nonprescription that patients take. Related to how people

think about medication is further pursuit of the implications of identifying medication by name, appearance, or purpose. The special concern is if there are transitions due to a change to a generic or in dosing and the ensuing challenges when refilling pill cases.

Related to this, a future study will seek to better understand the selection and use of pill cases over prescription bottles and the impact on remembering to take medication as prescribed, not taking extra doses when unsure if a dose was taken, and making refilling a case easier and more convenient. For both users of pill cases and prescription bottles, we aim to research how to support the development of more durable, disruption-proof triggers that function under routine and anomalous circumstances.

A final interest stemming from this study is the influence of home environments. We noted that participants who lived alone could develop a medication routine with substantial triggers with no limitations in terms of where they could place medication. An open question is whether this affords more flexibility to optimize the placement of a pill case or prescription bottle. Yet these same individuals lack someone in the home to remind them to take their medication, which for some provides an additional safety net. Another question is the influence of home size and the number of levels in a home in developing a strategy, in particular, how much proximity is needed between storage locations and heavily used locations in the home.

## Conclusions

The findings of this study provide important insights into the challenges older adults face in managing their medication and indicate opportunities to improve medication adherence in older adults. Our study participants formulated home medication management strategies without physician or pharmacist guidance, which represents an unexplored opportunity for improving adherence strategies through patient education provided by health care professionals during discussions with patients. Another opportunity is to reduce the extent to which trial and error is used to develop and refine medication management strategies through tailored guidance. Individualized guidance should include identifying more durable, disruption-proof triggers that function under routine and anomalous circumstances, including sufficient portability to accommodate travel. This qualitative study suggests that, in addition to supporting older adults who occasionally forget to take their medication, there are opportunities to improve adherence guidance to older adults when they are receiving their first long-term prescription or during a regimen change. It also suggests pathways for designing better adherence aids that integrate with established daily routines.

## Acknowledgments

Partial funding for open access was provided by Tufts University Hirsh Health Sciences Library's Open Access Fund. This research was funded in part by the Tufts University through the Springboard Program and in part by the National Center for Advancing Translational Sciences, National Institutes of Health (NIH; UL1TR002544). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors profusely thank the study participants who generously gave their time to describe their personal experiences in response to questions about their home medication management practices. The authors would like to thank Callan Moody, associate director, Osher Lifelong Learning Institute, who posted recruitment notices. The authors would also like to acknowledge and thank the students who were instrumental in developing this research, notably Ricardo Boschetti, Eden Shaveet, and Meera Singhal. The authors appreciate the feedback on this research and on this paper from In Baek. Finally, the authors thank the reviewers who provided feedback and suggestions that strengthened this paper.

## Data Availability

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

## Authors' Contributions

LG designed the study with assistance from MR and DW. LG conducted the interviews with assistance from MR, and both of them took notes. MR and DW conducted the thematic analysis of the interviews and independently replayed the interview recordings and used the transcripts to develop a list of preliminary codes, which were then reviewed by LG, MR, and DW and collated into potential themes until consensus was reached about the themes. MR and DW selected candidate quotes to illustrate the themes. Finally, LG, MR, and DW collated the themes into descriptive text and reviewed the chosen quotes to find exemplars for each theme. EM assisted in writing, reviewing, and editing the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 22 KB - ijmr_v13i1e53513_app1.docx ]

## References


1. Bouwman L, Eeltink CM, Visser O, Janssen JJ, Maaskant JM. Prevalence and associated factors of medication non-adherence in hematological-oncological patients in their home situation. BMC Cancer 2017 Nov 09;17(1):739 [FREE Full text] [doi: 10.1186/s12885-017-3735-1] [Medline: 29121889]
2. Benjamin RM. Medication adherence: helping patients take their medicines as directed. Public Health Rep 2012;127(1):2-3 [FREE Full text] [doi: 10.1177/003335491212700102] [Medline: 22298918]
3. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, et al. Medication compliance and persistence: terminology and definitions. Value Health 2008;11(1):44-47 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00213.x] [Medline: 18237359]
4. Adherence to long-term therapies: evidence for action. World Health Organization. 2003. URL: https://iris.who.int/handle/10665/42682 [accessed 2024-08-01]
5. Kvarnström K, Airaksinen M, Liira H. Barriers and facilitators to medication adherence: a qualitative study with general practitioners. BMJ Open 2018 Jan 23;8(1):e015332 [FREE Full text] [doi: 10.1136/bmjopen-2016-015332] [Medline: 29362241]
6. World population prospects 2019: highlights. United Nations Department of Economic and Social Affairs. 2019. URL: https://www.un.org/en/desa/world-population-prospects-2019-highlights [accessed 2024-04-24]
7. Kantor ED, Rehm CD, Haas JS, Chan AT, Giovannucci EL. Trends in prescription drug use among adults in the United States from 1999-2012. JAMA 2015 Nov 03;314(17):1818-1831 [FREE Full text] [doi: 10.1001/jama.2015.13766] [Medline: 26529160]
8. Horne R, Weinman J, Barber N, Elliott R, Morgan M. Concordance, adherence and compliance in medicine taking. National Institutes for Health and Care Research. 2005 Dec. URL: https://njl-admin.nihr.ac.uk/document/download/2027234 [accessed 2024-08-01]
9. Wieczorkiewicz SM, Kassamali Z, Danziger LH. Behind closed doors: medication storage and disposal in the home. Ann Pharmacother 2013 Mar 27;47(4):482-489. [doi: 10.1345/aph.1r706]
10. Rantz MJ, Zwygart-Stauffacher M, Popejoy L, Grando VT, Mehr DR, Hicks LL, et al. Nursing home care quality: a multidimensional theoretical model integrating the views of consumers and providers. J Nurs Care Qual 1999 Oct;14(1):16-37; quiz 85. [doi: 10.1097/00001786-199910000-00004] [Medline: 10575828]
11. Gandhi TK, Weingart SN, Borus J, Seger AC, Peterson J, Burdick E, et al. Adverse drug events in ambulatory care. N Engl J Med 2003 Apr 17;348(16):1556-1564. [doi: 10.1056/nejmsa020703]
12. Chan AH, Cooper V, Lycett H, Horne R. Practical barriers to medication adherence: what do current self- or observer-reported instruments assess? Front Pharmacol 2020 May 13;11:572 [FREE Full text] [doi: 10.3389/fphar.2020.00572] [Medline: 32477110]
13. Shaya FT, Eddington ND. Disruptive innovation in pharmacy: lessons from the Amazon frontier. JAMA Health Forum 2020 Jan 23;1(1):e200038 [FREE Full text] [doi: 10.1001/jamahealthforum.2020.0038] [Medline: 36218537]
14. Checchi KD, Huybrechts KF, Avorn J, Kesselheim AS. Electronic medication packaging devices and medication adherence: a systematic review. JAMA 2014 Sep 24;312(12):1237-1247 [FREE Full text] [doi: 10.1001/jama.2014.10059] [Medline: 25247520]
15. Ruane PJ, Parenti DM, Margolis DM, Shepp DH, Babinchak TJ, Van Kempen AS, et al. Compact quadruple therapy with the lamivudine/zidovudine combination tablet plus abacavir and efavirenz, followed by the lamivudine/zidovudine/abacavir triple nucleoside tablet plus efavirenz in treatment-naïve HIV-infected adults. HIV Clin Trials 2003;4(4):231-243. [doi: 10.1310/MM9W-BAU0-BT6Q-401B] [Medline: 12916008]
16. Conn VS, Ruppar TM, Chan KC, Dunbar-Jacob J, Pepper GA, De Geest S. Packaging interventions to increase medication adherence: systematic review and meta-analysis. Curr Med Res Opin 2015 Jan 04;31(1):145-160 [FREE Full text] [doi: 10.1185/03007995.2014.978939] [Medline: 25333709]
17. Notenboom K, Leufkens HG, Vromans H, Bouvy ML. Learning from patients: identifying design features of medicines that cause medication use problems. Int J Pharm 2017 Jan 30;517(1-2):128-134 [FREE Full text] [doi: 10.1016/j.ijpharm.2016.12.004] [Medline: 27931784]
18. Choudhry NK, Krumme AA, Ercole PM, Girdish C, Tong AY, Khan NF, et al. Effect of reminder devices on medication adherence: the REMIND randomized clinical trial. JAMA Intern Med 2017 May 01;177(5):624-631 [FREE Full text] [doi: 10.1001/jamainternmed.2016.9627] [Medline: 28241271]
19. Park JY, Li J, Howren A, Tsao NW, De Vera M. Mobile phone apps targeting medication adherence: quality assessment and content analysis of user reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]
20. Taitel M, Jiang J, Rudkin K, Ewing S, Duncan I. The impact of pharmacist face-to-face counseling to improve medication adherence among patients initiating statin therapy. Patient Prefer Adherence 2012;6:323-329 [FREE Full text] [doi: 10.2147/PPA.S29353] [Medline: 22563240]
21. Nguyen A, Uppal S, Pereira MM, Pluti A, Gualtieri L. MedHerent: improving medication adherence in older adults with contextually sensitive alerts through an application that adheres to you. Mayo Clinic Proc Digit Health 2024 Mar;2(1):1-7. [doi: 10.1016/j.mcpdig.2023.11.001]

22. Gualtieri L, Wang D, Baek I. Reevaluating the utility of time-based reminders to increase medication adherence in older adults. In: Proceedings of the 2024 SBM Annual Meeting. 2024 Presented at: SBM 2024; March 13-16, 2024; Philadelphia, PA. [doi: 10.1093/abm/kaae014]
23. Gualtieri L, Rosenbluth S. Can technology solve the problem of medication nonadherence? Mayo Clinic Proc Digit Health 2024 Mar;2(1):34-37. [doi: 10.1016/j.mcpdig.2023.11.008]
24. Wang D, Gualtieri L. Searching for medication adherence devices. In: Proceedings of the 27th Annual Meeting of ESPACOMP, the International Society for Medication Adherence. 2023 Presented at: ESPACOMP 2023; November 30-December 1, 2023; Budapest, Hungary.
25. Gualtieri L, Boschetti R, Rigby M, Shaveet E. Where are prescription medications stored: an investigation into home medication management practices. In: Proceedings of the 26th Annual Meeting of ESPACOMP, the International Society for Medication Adherence. 2022 Presented at: ESPACOMP 2022; November 17-19, 2022; Berlin, Germany.
26. Steiner JF. Rethinking adherence. Ann Intern Med 2012 Oct 16;157(8):580-585. [doi: 10.7326/0003-4819-157-8-201210160-00013] [Medline: 23070491]
27. Keller J, Kwasnicka D, Klaiber P, Sichert L, Lally P, Fleig L. Habit formation following routine-based versus time-based cue planning: a randomized controlled trial. Br J Health Psychol 2021 Sep;26(3):807-824. [doi: 10.1111/bjhp.12504] [Medline: 33405284]
28. Arlinghaus KR, Johnston CA. The importance of creating habits and routine. Am J Lifestyle Med 2019 Dec 29;13(2):142-144 [FREE Full text] [doi: 10.1177/1559827618818044] [Medline: 30800018]
29. Gualtieri L, Shaveet E, Estime B, Patel A. The role of home medication storage location in increasing medication adherence for middle-aged and older adults. Front Digit Health 2022 Nov 4;4:999981 [FREE Full text] [doi: 10.3389/fdgth.2022.999981] [Medline: 36405415]
30. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, et al. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
31. Osher Lifelong Learning Institute at Tufts University. Tufts University. URL: https://universitycollege.tufts.edu/lifelong-learning [accessed 2023-10-08]
32. Smith JA. Qualitative Psychology: A Practical Guide to Research Methods. Thousand Oaks, CA: SAGE Publications; 2015.
33. Schwartz JK. Pillbox use, satisfaction, and effectiveness among persons with chronic health conditions. Assist Technol 2017 Sep 30;29(4):181-187. [doi: 10.1080/10400435.2016.1219884] [Medline: 27689861]
34. Mickelson RS, Holden RJ. Medication adherence: staying within the boundaries of safety. Ergonomics 2018 Jan 10;61(1):82-103 [FREE Full text] [doi: 10.1080/00140139.2017.1301574] [Medline: 28394204]


## Abbreviations

**OLLI:** Osher Lifelong Learning Institute


*Edited by T de Azevedo Cardoso; submitted 09.10.23; peer-reviewed by B Pouls, R Marcilly, B Estime; comments to author 03.04.24; revised version received 26.04.24; accepted 20.06.24; published 13.08.24.*

*Please cite as:*
*Gualtieri L, Rigby M, Wang D, Mann E*
*Medication Management Strategies to Support Medication Adherence: Interview Study With Older Adults*
*Interact J Med Res 2024;13:e53513*
*URL: https://www.i-jmr.org/2024/1/e53513*
*doi:10.2196/53513*
*PMID:39137021*

Original Paper

# Portuguese Version of the Oral Frailty Index-8: Instrument Validation Study

Laura Corrêa[1], MSc; André Júdice[1], DMD; Robson Scoz[1], DPhil; Vanessa Machado[1], DPhil, DMD; José João Mendes[1], DPhil, DMD; Luís Proença[1], DPhil; João Botelho[1], DPhil, DMD; Luciano Ferreira[1], DPhil

Egas Moniz School of Health and Science, Almada, Portugal

**Corresponding Author:**
João Botelho, DPhil, DMD
Egas Moniz School of Health and Science
Campus Universitário, Quinta da Granja, Monte de Caparica
Almada, 2829-511
Portugal
Phone: 351 969848394
Email: jbotelho@egasmoniz.edu.pt

## Abstract

**Background:** The concept of oral frailty has gained scientific and clinical relevance in recent years, and early detection can facilitate timely intervention to manage its progression. The Oral Frailty Index-8 (OFI-8) was developed to assess community-dwelling older adults at risk for oral frailty.

**Objective:** This study aims to investigate the psychometric validity of the OFI-8 in the Portuguese population, named the Portuguese version of the OFI-8 (OFI-8-PT), which may serve as a reference for future studies related to longevity and oral function.

**Methods:** This study included 2 main phases, involving patients aged 60 years or older, Portuguese speakers, and those who consented to participate in the study. First, the researchers translated and cross-culturally adapted the original questionnaire to make it suitable for native Portuguese speakers. The translated tool was then assessed for psychometric validation, which consisted of test-retest reliability, internal consistency, construct validity, and sex invariance measurement.

**Results:** A total of 159 older adults participated in the baseline survey, with almost equal numbers of male (n=79, 49.7%) and female participants (n=80, 50.3%). The OFI-8-PT demonstrated good reliability (Cronbach $\alpha$=0.95) and construct validity (goodness-of-fit index=0.96; comparative fit index=0.85; and root mean square error of approximation=0.05, 90% CI 0.00-0.09). The study found sex invariance, indicating that the OFI-8-PT is equally valid for male and female participants, and the tested-retest reliability of the OFI-8-PT was good, indicating consistent results over time.

**Conclusions:** The OFI-8-PT showed psychometric validity and good reliability to be used in the Portuguese population.

***(Interact J Med Res 2024;13:e49975)*** doi:10.2196/49975

**KEYWORDS**

oral frailty; oral health; functional disability; frailty; aging; dentistry; confirmatory factor analysis; psychometric validity; questionnaire development

## Introduction

Contemporary societies are grappling with an unprecedented aging phenomenon, propelled by the evolving age distribution of the global population [1,2], largely owing to remarkable gains in life expectancy [3]. The increasing longevity—considered a testament to medical advancements, improved access to treatments, and an emphasis on preventive care including oral health maintenance [1,4,5]—presents a notable societal achievement. However, amid this transition emerges the challenge of frailty—a condition marked by heightened vulnerability attributed to age-related declines in physiological reserves [6,7]. Frailty, often exacerbated by a prolonged lifespan, poses distinct health and societal complexities [2], underscoring the multifaceted nature of aging in contemporary times.

Within this spectrum, oral frailty was introduced in 2013 in Japan and has gained increased attention in recent years [8,9]. Oral frailty is defined as a decline in oral health and functional capacity, with broader implications for overall health and quality of life [8,9]. Tanaka et al [10] developed an 8-item

questionnaire, referred to as the Oral Frailty Index-8 (OFI-8), for the purpose of surveying community-dwelling older adults who are at risk of oral frailty. Evidence suggests that early identification and intervention can slow the progression of oral frailty and may even prevent its onset [11,12]. This has led to a growing international interest in identifying the effects of aging on oral health and appropriate strategies for its prevention [11,13]. For this reason, oral frailty is currently considered one of the main determinants of oral health and dentistry policies [9].

Maintaining good oral function can then be one of the keys to increasing longevity, but the evidence is limited and has not been studied in detail [8]. Thus, additional high-quality studies are needed, and it is extremely important to increase awareness of oral frailty and provide appropriate literature related to this concept to promote healthy aging in the future [5,10].

Therefore, considering the potential impact of the OFI-8 and the need for further validation worldwide, the purpose of this research is to serve as a reference for future studies related to longevity and focused on oral function by translating and exploring the psychometric validity of OFI-8 in the Portuguese population, named the Portuguese version of the OFI-8 (OFI-8-PT).

## Methods

### Design, Setting, and Participants

This was a cross-sectional study that aimed to translate and validate the OFI-8-PT in a sample of patients aged 60 years older from the Egas Moniz Dental Clinic (Almada, Portugal).

To be eligible, participants needed to be aged 60 years or older; be fluent in Portuguese; and consent to participate. The recruitment was carried out between December 30, 2022, and February 17, 2023, which corresponds to a total period of approximately 7 weeks. The recruitment procedure and sampling method consisted of a consecutive convenience sample of participants eligible for participation that presented at the Egas Moniz Dental Clinic for a first dental appointment.

### Sample Size Calculation

To achieve the proposed aim, we calculated a minimum sample size based on the rules of thumb from Terwee et al [14] (between 4 and 10 participants per variable), with a minimum of 10 individuals per questionnaire item. As such, considering an 8-item questionnaire with a dichotomous nature, we estimated a minimum number of 80 participants to ensure the stability of the variance-covariance matrix when performing confirmatory factor analysis (CFA).

### Translation and Cross-Cultural Adaptation of the OFI-8

The original OFI-8 questionnaire measures the oral frailty using 8 dichotomous items, representing the number of natural teeth, masticatory performance, maximum tongue pressure, articulatory oral motor skill, subjective difficulties in eating tough food, and subjective swallowing difficulties.

Each question is assigned a score that is then added together to obtain the total score of the questionnaire, resulting in a numerical scale that ranges from 0 to 11 points; higher scores indicate a greater degree of oral frailty, which in turn has other implications previously described, such as physical frailty, dependence, and mortality [5]. For questions 1, 2, and 3, a response of “Yes” was allocated 2 points, while for questions 4 and 5, a response of “Yes” was allocated 1 point. In contrast, a “No” response was awarded 0 points. Conversely, for questions 6, 7, and 8, a response of “No” was given 1 point, while a response of “Yes” was not awarded any points [10].

Two independent, bilingual individuals fluent in both Portuguese and English translated the original English version of the OFI-8 into Portuguese (VM and JB) in a “double-blind” process. Because no disagreements were recorded, a final version of the questionnaire was named OFI-8-PT. After back-translation to English and comparing with the original to identify discrepancies, no ambiguous wording or cultural nuances were identified. The semantic and conceptual equivalence was attested (Table 1).

**Table 1.** Original version and Portuguese translation of the Oral Frailty Index-8 (OFI-8).

| Item | Original (English) | Translation (Portuguese) |
|---|---|---|
| 1 | Do you have any difficulties eating tough foods compared to 6 months ago? | Em comparação com há 6 meses, sentiu dificuldades em comer alimentos duros? |
| 2 | Have you choked on your tea or soup recently? | Recentemente, engasgou-se com chá ou sopa? |
| 3 | Do you use dentures? | Usa prótese dentária? |
| 4 | Do you often have a dry mouth? | Costuma ter a boca seca? |
| 5 | Do you go out less frequently than you did last year? | Sai de casa com menos frequência do que no ano passado? |
| 6 | Can you eat hard foods like squid jerky or pickled radish? | Consegue comer alimentos duros como carne seca ou nozes? |
| 7 | How many times do you brush your teeth in a day? (2 or more times/day) | Quantas vezes escova os dentes por dia (2 ou mais vezes/dia) |
| 8 | Do you visit a dental clinic at least annually? | Visita o Médico Dentista pelo menos uma vez por ano? |

### Variables

The dependent variable in this study was the older adults' self-perception of their own oral health (oral frailty score), which occurs as a result of the independent variable. The way in which participants define their oral health condition is used as an indicator for the oral frailty score.

Some parameters of the participants in this study, such as age and sex, were considered independent variables of the study. These variables can be used to determine if there is any relationship with the oral frailty score of the individuals.

### Statistical Analyses

The statistical analysis was conducted in the R *plyr* package (R Foundation for Statistical Computing).

#### *Reliability*

The OFI-8-PT was pilot-tested on a random sample of 10% of the total sample required for validation (and this subsample was not part of the validation stage). The eligibility criteria of this pilot test were the same as the validation. We used the Kuder-Richardson formula 20 to test the internal consistency of each item, considering its dichotomous nature [15].

#### *Validity*

Using the *lavaan* package, we performed the CFA to compute the factorial loads and model fitness. To compute the model, we used the maximum likelihood method, with the differences between models being explored through chi-square ($\chi^2$) and likelihood ratio tests. To test the fitness of CFA, we used the $\chi^2/df$ ratio (good adjustment with values <2) [16], the root mean squared error of approximation (RMSEA; a good model adjustment considered for values between 0.05% and 0.10%; 90% CI) [17], the confirmatory fit index (CFI; a cutoff criterion of ≥0.90 indicates a good fit) [18], goodness-of-fit index (GFI; values of 0.90 or greater indicate well-fitting models) [19], and normed fit index (a cutoff criterion of 0.90) [20].

Then, to make sure that the change of the state variables indicates state variability and not trait change, such as sex, we analyzed the sex invariance of the OFI-8-PT. To do this, we estimated 4 consecutive models: (1) unconstrained model; (2) factor loading–constrained model (Model 1); (3) factor loading– and structural covariance–constrained model (Model 2); and (4) factor loading–, structural covariance–, and measurement residual–constrained model (Model 3). Then, we were able to obtain the Δ values for CFI (ΔCFI) and chi-square ($\Delta\chi^2$). We defined a cutoff point of ΔCFI<0.01 as the presence of invariance [18,21] and $\Delta\chi^2$=0.095 as invariance between the models [22,23]. In addition, we explored the relationships between the OFI-8-PT items to confirm instrument dimensionality via the Spearman rank correlation coefficient (ρ) and the effect of the previously mentioned 2 confounding variables, sex and age, through the Spearman correlation with the overall OFI-8-PT score. The level of statistical significance was set at 5% in all inferential analyses.

### Ethical Considerations

The data collection was performed electronically, ensuring the confidentiality and anonymity of the participants' data through a cloud-based system [24]. The study was conducted in accordance with the regulations applicable to research and was approved by the Ethics Committee of the Egas Moniz (ID: 1140/2022). Informed consent was obtained from each participant before their inclusion in the study. Participants entered into this research voluntarily, fully aware of their right to withdraw from the study at any point without the need for justification. The informed consent process ensured that participants had a clear understanding of the study's purpose, procedures, and the voluntary nature of their involvement. Patients received compensation in the form of free diagnosis and treatment without costs.

## Results

### Participant Characteristics

A total of 159 participants completed the OFI-8-PT, with an average age of 73.9 (SD 9.4; range 60-99) years (mean 74.8, SD 9.7 years and mean 73.1, SD 9.0 years for female and male participants, respectively), and the group was equally composed by female and male participants (n=80, 50.3% female vs n=79, 49.7% male).

### Test-Retest Reliability

To test the reliability of the OFI-8-PT, 16 participants answered the translated tool 2 times (baseline and 2 weeks after). Of these participants, 6 (38%) were female and 10 (62%) were male, with similar age intervals (female: mean 70.8, SD 4.7 years vs male: mean 71.7, SD 9.2 years). The median total score of the instrument was 3 (range 0-7).

The overall internal consistency through Kuder-Richardson formula 20 was 0.69, thus considered to be a good reliability score. Nominally, each question had excellent reliability (Multimedia Appendix 1).

The OFI-8-PT exhibited an adequate reliability (with a Cronbach α coefficient of 0.95) and an adequate psychometric feature.

### Construct Validity

The CFA analysis attested the unifactorial structure of the OFI-8-PT. The first-order unifactorial model resulted in an adequate model fit (GFI=0.96; CFI=0.85; and RMSEA=0.05, 90% CI 0.00-0.09; Table 2).

**Table 2.** Model fit indices in the unifactorial model and configurational invariance by sex.

| Model | $\chi^2$ | *df* | *P* value | $\chi^2/df$ | CFI[a] | GFI[b] | RMSEA[c] (90% CI) | ΔCFI | $\Delta\chi^2$ | Δ*df* |
|---|---|---|---|---|---|---|---|---|---|---|
| **Unifactorial** | 26.5 | 20 | <.001 | 1.33 | 0.85 | 0.96 | 0.05 (0.00-0.09) | —[d] | — | — |
| **Measurement invariance across sex** | | | | | | | | | | |
| Unconstrained | 43.8 | 40 | — | 1.10 | 0.90 | 0.96 | 0.04 (0.00-0.09) | — | — | — |
| 1 | 48.6 | 47 | <.001 | 1.03 | 1.00 | 0.96 | 0.00 (0.00-0.07) | 0.10 | 4.8 | 7 |
| 2 | 56.0 | 54 | <.001 | 1.04 | 0.95 | 0.99 | 0.02 (0.00-0.08) | 0.05 | 7.4 | 7 |
| 3 | 56.0 | 54 | <.001 | 1.04 | 0.95 | 0.99 | 0.02 (0.00-0.08) | 0.00 | 0.0 | 0 |

[a]CFI: comparative fit index.

[b]GFI: goodness-of-fit index.

[c]RMSEA: root mean square error of approximation.

[d]Not applicable.

A multigroup CFA assessed the invariance across sex groups in the OFI-8-PT (Table 1). Accordingly, there was invariance for sex groups based on the comparisons of CFI, $\chi^2$, and *df* values across the unconstrained and constrained models studied.

In order to learn the agreement level of OFI-8-PT components, we analyzed the correlation between the items using Spearman rank correlation coefficient. There was a low number of significant correlations (4/28, 14% of all correlations; Table 3).

**Table 3.** Correlation between Portuguese version of the Oral Frailty Index-8 item scores.

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **1** | | | | | | | | |
| *r* | 1 | 0.103 | 0.166 | –0.04 | 0.215 | 0.331 | 0.112 | 0.066 |
| *P* value | —[a] | >.99 | .38 | .61 | .29 | .01 | .49 | .49 |
| **2** | | | | | | | | |
| *r* | 0.103 | 1 | 0.175 | 0.044 | 0.075 | 0.07 | 0.012 | 0.133 |
| *P* value | >.99 | — | .21 | .91 | .29 | .49 | .16 | .74 |
| **3** | | | | | | | | |
| *r* | 0.166 | 0.175 | 1 | 0.117 | 0.024 | 0.205 | –0.105 | –0.037 |
| *P* value | .38 | .21 | — | .14 | .09 | .23 | .09 | .047 |
| **4** | | | | | | | | |
| *r* | –0.04 | 0.044 | 0.117 | 1 | 0.086 | 0.102 | –0.042 | –0.005 |
| *P* value | .61 | .91 | .14 | — | .14 | .91 | .12 | .16 |
| **5** | | | | | | | | |
| *r* | 0.215 | 0.075 | 0.024 | 0.086 | 1 | 0.142 | 0.187 | 0.145 |
| *P* value | .29 | .29 | .09 | .14 | — | .78 | .03 | .18 |
| **6** | | | | | | | | |
| *r* | 0.331 | 0.07 | 0.205 | 0.102 | 0.142 | 1 | 0.181 | –0.009 |
| *P* value | .01 | .49 | .23 | .91 | .78 | — | .78 | .047 |
| **7** | | | | | | | | |
| *r* | 0.112 | 0.012 | –0.105 | –0.042 | 0.187 | 0.181 | 1 | 0.123 |
| *P* value | .49 | .16 | .09 | .12 | .03 | .78 | — | .23 |
| **8** | | | | | | | | |
| *r* | 0.066 | 0.133 | –0.037 | –0.005 | 0.145 | –0.009 | 0.123 | 1 |
| *P* value | .49 | .74 | .047 | .16 | .18 | .047 | .23 | — |

[a]Not applicable.

### Confounding Variables

When correlating the overall OFI-8-PT score, age was significantly correlated (ρ=0.259; *P*<.001), whereas sex was not (ρ=–0.035; *P*=.66), confirming as well the absence of sex invariance.

## Discussion

### Principal Findings

This study demonstrated the psychometric validity of the OFI-8-PT to depict individuals at risk of oral frailty and functional disability in this Portuguese sample. In addition, we found an adequate internal consistency and reliability of this questionnaire.

The assessment of oral frailty in older adult individuals through a comprehensive scale holds great clinical significance due to its potential influence on geriatric care and the overall health outcomes of these individuals. Older adults typically face a variety of oral health challenges, such as tooth loss, periodontal disease, and dry mouth, which can significantly impact their quality of life and general well-being [25,26]. Despite the prevalence of oral health problems in the older adult population, there is a scarcity of specialized tools designed to assess oral frailty, which encompasses the functional and structural vulnerabilities of the oral cavity [27,28].

Expanding the usability of the OFI-8 to other languages could help health care professionals detect, intervene, and track oral health in older adults, potentially improving their quality of life through preventive care and tailored hygiene. Moreover, incorporating oral frailty assessment into routine geriatric assessments can facilitate comprehensive care planning and interdisciplinary collaboration among health care providers [29]. By addressing the clinical need for a validated scale to assess oral frailty in older adults, this study contributes to enhancing the holistic approach to geriatric health care and underscores the importance of oral health in promoting healthy aging.

The GFI indicated that the observed covariance matrix of the data fitted the covariance matrix implied by the model, revealing a very good fit. The CFI, however, suggested a reasonable fit with room for improvement when compared to a null model. The RMSEA also revealed a good fit with a value of 0.05. These results support the validity of the OFI-8-PT, although they cannot be directly compared with the original study by Tanaka et al [5] due to differences in methodology. Additionally, this is the first study to validate the OFI-8 in a language other than Japanese, which may contribute to future comparability with other validations and cultural adaptations.

With the introduction of the OFI-8-PT, health professionals and researchers will have a valuable tool for assessing oral frailty in older adults in Portugal. This standardized and validated questionnaire will enable early detection of oral frailty, allowing for targeted interventions and preventive strategies. With the current aging population, understanding and managing oral frailty is becoming increasingly important to promote healthy aging and improve overall well-being. By identifying individuals at risk, the OFI-8-PT can facilitate appropriate oral health services, ultimately contributing to improved oral health outcomes and quality of life.

The OFI-8-PT questionnaire offers several advantages that make it a valuable tool for assessing oral frailty in older adults in Portugal. Its standardized and validated nature ensures reliable and consistent assessment, while its ease of administration and scoring facilitates its incorporation into routine assessments. Implementation of the OFI-8-PT also allows for data collection on oral frailty in the Portuguese population, contributing to a better understanding of its prevalence and impact on health outcomes. However, it is important to acknowledge potential drawbacks, such as its primary focus on oral health aspects and its reliance on self-reporting, which may introduce bias or inaccuracy. Addressing these limitations and ensuring proper training and awareness of participants will be critical to maximizing the benefits of the OFI-8-PT questionnaire.

### Strengths and Limitations

Overall, the study provides valuable insights into the psychometric validity of the OFI-8-PT questionnaire in the Portuguese population. However, there are some important limitations worth discussing that should be considered when interpreting the results.

First, the sample size used in the study was relatively small, which may limit the generalizability of the findings to a larger population. Yet, this shortcoming is limited as the OFI-8 is a 1-factor scale and we followed a previous validated strategy. Additionally, the study relied solely on self-report measures, based on self-perception of oral health, which could result in bias or social desirability effects. Furthermore, the study was conducted in a specific population from a university clinic, consisting solely of patients who visited the dental clinic, and therefore, findings may not be generalizable to the entire Portuguese population or other populations in different geographical locations. Nevertheless, these participants live in the largest metropolitan area in Portugal and may provide a considerable percentage of representation. Moreover, the presence of confounding variables, such as chronic diseases and medications, could influence the oral frailty score of individuals. Although these variables were acknowledged, the study did not extensively analyze their impact or control for their effects, and the study could not assess longitudinal effects due to its design.

Accordingly, conducting further research with larger and more diverse samples is recommended to address these limitations and strengthen the validity of the findings.

### Conclusion

In conclusion, the OFI-8-PT was shown to have psychometric validity and reliability in the population under study, providing a snapshot of its oral frailty and functional disability. This consistency and ease of use may contribute to better screening and monitoring oral frailty prevalence and its impact on health outcomes. It may have an impact on future public health strategies to address oral frailty.

## Acknowledgments

This study is part of the project “Biological Signs” funded by Egas Moniz School of Health and Science (Egas Moniz Centre for Interdisciplinary Research [CiiEM] research grant). The authors acknowledge the volunteer participants from the study carried out at the Egas Moniz Dental Clinic. This study is supported by an institutional grant from *Fundação para a Ciência e Tecnologia* (10.54499/UIDB/04585/2020).

## Data Availability

The datasets generated during and analyzed during this study are available in a Zenodo repository [30].

## Authors' Contributions

LF, AJ, VM, LP, and JB conceptualized the study. LC and JB were responsible for data curation. JB conducted the formal analysis. AJ, RS, VM, JJM, LP, JB, and LF contributed to the investigation. VM, JB, and LF developed the methodology. LF managed the project administration. JJM, LP, JB, and LF provided resources. JB and LF supervised the project. JB performed the validation. All authors contributed to writing the original draft and reviewing and editing the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Reliability for each question of the Portuguese version of the Oral Frailty Index-8.
[DOCX File , 18 KB - ijmr_v13i1e49975_app1.docx ]

## References

1. United Nations Department for Economic and Social Affairs. World Population Prospects 2022: Summary of Results. New York, NY: United Nations; 2023.
2. United Nations Department for Economic and Social Affairs. World Population Prospects 2019: Highlights. New York, NY: United Nations; 2019.
3. Life expectancy across EU regions in 2020. Eurostat. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20220427-1 [accessed 2023-02-12]
4. Singh S, Bajorek B. Defining 'elderly' in clinical practice guidelines for pharmacotherapy. Pharm Pract (Granada) 2014;12(4):489 [FREE Full text] [doi: 10.4321/s1886-36552014000400007] [Medline: 25580172]
5. Tanaka T, Takahashi K, Hirano H, Kikutani T, Watanabe Y, Ohara Y, et al. Oral frailty as a risk factor for physical frailty and mortality in community-dwelling elderly. J Gerontol A Biol Sci Med Sci 2018;73(12):1661-1667. [doi: 10.1093/gerona/glx225] [Medline: 29161342]
6. Bandeen-Roche K, Xue Q, Ferrucci L, Walston J, Guralnik JM, Chaves P, et al. Phenotype of frailty: characterization in the women's health and aging studies. J Gerontol A Biol Sci Med Sci 2006;61(3):262-266. [doi: 10.1093/gerona/61.3.262] [Medline: 16567375]
7. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
8. Iwasaki M, Hirano H. Decline in oral function and its management. Int Dent J 2022;72(4S):S12-S20 [FREE Full text] [doi: 10.1016/j.identj.2022.06.010] [Medline: 36031320]
9. Watanabe Y, Okada K, Kondo M, Matsushita T, Nakazawa S, Yamazaki Y. Oral health for achieving longevity. Geriatr Gerontol Int 2020;20(6):526-538. [doi: 10.1111/ggi.13921] [Medline: 32307825]
10. Tanaka T, Hirano H, Ohara Y, Nishimoto M, Iijima K. Oral Frailty Index-8 in the risk assessment of new-onset oral frailty and functional disability among community-dwelling older adults. Arch Gerontol Geriatr 2021;94:104340. [doi: 10.1016/j.archger.2021.104340] [Medline: 33529863]
11. Nomura Y, Ishii Y, Chiba Y, Suzuki S, Suzuki A, Suzuki S, et al. Structure and validity of questionnaire for oral frail screening. Healthcare (Basel) 2021;9(1):45 [FREE Full text] [doi: 10.3390/healthcare9010045] [Medline: 33466528]
12. Dros C, Sealy MJ, Krijnen WP, Weening-Verbree LF, Hobbelen H, Jager-Wittenaar H. Oral health and frailty in community-dwelling older adults in the Northern Netherlands: a cross-sectional study. Int J Environ Res Public Health 2022;19(13):7654 [FREE Full text] [doi: 10.3390/ijerph19137654] [Medline: 35805314]
13. Minakuchi S, Tsuga K, Ikebe K, Ueda T, Tamura F, Nagao K, et al. Oral hypofunction in the older population: position paper of the Japanese society of gerodontology in 2016. Gerodontology 2018;35(4):317-324. [doi: 10.1111/ger.12347] [Medline: 29882364]
14. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]

15. Kuder GF, Richardson MW. The theory of the estimation of test reliability. Psychometrika 1937;2(3):151-160. [doi: 10.1007/bf02288391]
16. Tabachnick B, Fidell L, Ullman J. Using Multivariate Statistics. Seventh Edition. New York, NY: Pearson; 2019.
17. Steiger JH. Understanding the limitations of global fit assessment in structural equation modeling. Personality and Individual Differences 2007;42(5):893-898. [doi: 10.1016/j.paid.2006.09.017]
18. Byrne B. Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming, Third Edition. New York, NY: Routledge, Taylor & Francis Group; 2016.
19. Shevlin M, Miles J. Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. Personality and Individual Differences 1998;25(1):85-90. [doi: 10.1016/S0191-8869(98)00055-5]
20. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
21. Dimitrov DM. Testing for factorial invariance in the context of construct validation. Measurement and Evaluation in Counseling and Development 2017;43(2):121-149. [doi: 10.1177/0748175610373459]
22. Bollen K. Structural Equations with Latent Variables. New York, NY: Wiley; 1989.
23. Satorra A, Bentler PM. Ensuring Positiveness of the scaled difference chi-square test statistic. Psychometrika 2009;75(2):243-248. [doi: 10.1007/s11336-009-9135-y]
24. Mondal H, Mondal S, Ghosal T, Mondal S. Using Google forms for medical survey: a technical note. IJCEP 2019;5(4):216-218. [doi: 10.5530/ijcep.2018.5.4.26]
25. Petersen PE, Yamamoto T. Improving the oral health of older people: the approach of the WHO global oral health programme. Community Dent Oral Epidemiol 2005;33(2):81-92. [doi: 10.1111/j.1600-0528.2004.00219.x] [Medline: 15725170]
26. Glick M, Williams DM, Kleinman DV, Vujicic M, Watt RG, Weyant RJ. A new definition for oral health developed by the FDI world dental federation opens the door to a universal definition of oral health. J Public Health Dent 2017;77(1):3-5. [doi: 10.1111/jphd.12213] [Medline: 28276588]
27. Thomson WM, Chalmers JM, Spencer AJ, Williams SM. The Xerostomia inventory: a multi-item approach to measuring dry mouth. Community Dent Health 1999;16(1):12-17. [Medline: 10697349]
28. Gill TM, Gahbauer EA, Han L, Allore HG. Trajectories of disability in the last year of life. N Engl J Med 2010;362(13):1173-1180 [FREE Full text] [doi: 10.1056/NEJMoa0909087] [Medline: 20357280]
29. Inouye SK, Studenski S, Tinetti ME, Kuchel GA. Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept. J Am Geriatr Soc 2007;55(5):780-791 [FREE Full text] [doi: 10.1111/j.1532-5415.2007.01156.x] [Medline: 17493201]
30. Botelho J. Datasets of the OFI-8-PT test-retest and validation. Zenodo. 2024 Sep 17. URL: https://zenodo.org/records/13771746 [accessed 2024-10-03]


## Abbreviations

**CFA:** confirmatory factor analysis
**CFI:** comparative fit index
**GFI:** goodness-of-fit index
**OFI-8:** Oral Frailty Index-8
**OFI-8-PT:** Portuguese version of the Oral Frailty Index-8
**RMSEA:** root mean square error of approximation


*Edited by T de Azevedo Cardoso; submitted 15.06.23; peer-reviewed by M Pang, M Sarhan; comments to author 19.12.23; revised version received 12.02.24; accepted 11.07.24; published 28.10.24.*

*Please cite as:*
*Corrêa L, Júdice A, Scoz R, Machado V, Mendes JJ, Proença L, Botelho J, Ferreira L*
*Portuguese Version of the Oral Frailty Index-8: Instrument Validation Study*
*Interact J Med Res 2024;13:e49975*
*URL: https://www.i-jmr.org/2024/1/e49975*
*doi:10.2196/49975*
*PMID:39466299*

Original Paper

# Patient Profile and Cost Savings of Long-Term Care in a Spanish Hospital: Retrospective Observational Study

José Joaquín Mira[1,2], MPH, PhD; Daniel García-Torres[2], MSc; María del Mar Bonell-Guerrero[3], RN; Ana Isabel Cáceres-Sevilla[3], RN; Martina Ramirez-Sanz[3], RN; Rosa Martínez-Lleo[3], RN; Concepción Carratalá[1], PhD

[1]Health Psychology Department, Universidad Miguel Hernández, Elche, Spain

[2]Foundation for the Promotion of Health and Biomedical Research of the Valencian Community (FISABIO), Alicante, Spain

[3]Hospital Padre Jofre, Valencia, Spain

**Corresponding Author:**
José Joaquín Mira, MPH, PhD
Health Psychology Department
Universidad Miguel Hernández
Universidad Avenue
Elche, 03202
Spain
Phone: 34 966658984
Email: jose.mira@umh.es

## Abstract

**Background:** Long-term care hospitals have been considered an efficient response to the health care needs of an increasingly aging population. These centers are expected to contribute to better hospital bed management and more personalized care for patients needing continuous care. The evaluation of their outcomes is necessary after a sufficient period to assess their impact. Hospitals for Acute and Chronic Long-Term Extended Stay (HACLES) emerged in Spain in the late 20th century as a response to the aging population and the increase in chronic diseases.

**Objective:** This study aimed to analyze the profile of patients treated in a HACLES, particularly analyzing gender differences, and evaluate the cost savings associated with using these centers.

**Methods:** A retrospective study was conducted based on data from patients 65 years old or older admitted to a HACLES between 2022 and 2023. Gender, age, household cohabitation data, diagnosis and comorbidity, daily medication intake, and degree of dependency were obtained to describe the profile of patients who attended the HACLES. Data coded in SIA-Abucasis (version 37.00.03; Consellería Sanitat, Generalitat Valenciana; a digital medical record system used in the Valencian region) were reviewed, and descriptive statistics and comparison tests were used. The direct cost savings of HACLES admissions were calculated by comparing the daily cost of a general hospital bed with that of a HACLES bed.

**Results:** Data from 123 patients with a mean age of 77 years were analyzed. Most (n=81, 65.9%) had a cohabiting family member as their primary caregiver. Palliative care was the most frequent reason for admission (n=75, 61%). The mortality rate (odds ratio [OR] 61.8, 95% CI 53.2-70.5) was similar between men and women (OR 54.1, 95% CI 47.8-71.5 vs OR 59.7, 95% CI 42.2-66.0; *P*=.23). The cognitive assessment, using the Pfeiffer scale, improved at discharge (mean 3.2, SD 3.2 vs mean 2.5, SD 3.1; *P*=.003). The length of stay was significantly larger for patients who returned home compared with patients discharged to other facilities (mean 89.8, SD 58.2 versus mean 33.1, SD 43.1 days; *P*<.001). The direct cost savings were estimated at US $42,614,846 per 1000 admissions.

**Conclusions:** Patients typically treated in HACLES are older, with a high level of cognitive impairment and physical dependency, and a significant proportion are in palliative care, highlighting the importance of adapting care to the individual needs of the admitted patients. The HACLES model contributes to the sustainability of the public health system.

***(Interact J Med Res 2024;13:e64248)*** doi:10.2196/64248

**KEYWORDS**

chronicity; length of stay; hospital; chronic; long-term care; demographics; gerontology; Hospitals for Acute and Chronic Long-Term Extended Stay; HACLES; healthcare economics; cost savings

## Introduction

In all high-income countries, long-term care hospitals have been considered an efficient response to the health care needs of an increasingly aging population that have multiple chronic conditions simultaneously [1]. However, the care model is being reviewed, and new ideas are emerging to leverage the potential of these health care resources [2].

In Spain, Hospitals for Acute and Chronic Long-Term Extended Stay (HACLES) were initially designed to free up beds in general hospitals occupied by patients with chronic conditions who did not require an acute care but needed a suitable hospital environment [3]. These hospitals continue to focus on managing complex medical needs that extend beyond short-term acute care settings, with an emphasis on long-term care and support for patients with chronic illnesses, disabilities, or conditions requiring prolonged medical treatment.

In the mid-90s, HACLES were established in the Valencian Community as a response to the need for specialized care for patients with chronic conditions and those requiring long-term care. These centers integrated medical, nursing, rehabilitation, psychological, and social support [4]. By the early 21st century, this model was consolidated with investments to make them more accessible and comfortable for prolonged stays, adapting to the profiles and needs of patients with chronic conditions, including those requiring palliative care. These centers are expected to contribute to better hospital bed management and more personalized care for patients needing continuous care.

The HACLES model, as is happening in other countries [5] and having gone through its period of implosion and consolidation, requires the development of long-term strategies that respond to the demographic and technological changes in our society. This study aims to describe the profile of patients receiving care in a HACLES, particularly analyzing gender differences, and evaluate the cost savings associated with using these centers.

## Methods

### Overview

A retrospective observational study was conducted based on the review of data coded in SIA-Abucasis (the digital clinical history system used in the Valencian Community) from a systematic sample of patients >65 years old admitted to the HACLES. The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines were used to describe this study (Multimedia Appendix 1).

### Patient Eligible Criteria

Data were recorded from all patients admitted over a period 2 years, between January 2022 and December 2023. A blinded registration system was used for the research team. A web-based platform was developed to ensure data quality, accessed through a personalized key, respecting data anonymization.

### Sample Size

Considering the annual number of admissions (N=294), an $\alpha$ risk of .05, and a precision of 5%, the sample size was estimated at 168 medical records. This calculation was made using the formula for sample size in finite populations.

### Sample Selection

Among the patients randomly assigned to each reviewer, a total of 30 cases were selected through simple randomization (k=3).

The reviewers included 4 nurses (all women). The reviewers were trained in the review procedure and in using a data registration tool, ensuring uniformity in the interpretation of the study protocol. A call center was available during the field study to resolve any issues related to data entry on the web platform.

### Study Variables

These included binary variables such as gender; continuous variables such as age, polypharmacy (daily medication intake), average stay, and, when available, Barthel [6], Pfeiffer [7], and Gijón [8] scales scores at admission and discharge; and nominal variables such as household cohabitation data, primary diagnosis, comorbidity (based on the *International Classification of Diseases, 10th revision* [*ICD-10*]), reason for admission, degree of dependency, residence at the time of admission, destination upon discharge (home, institution, death, or others), and who the primary caregiver is.

### Cost Analysis

The direct cost savings represented by admission to a HACLES were estimated by the difference between the daily cost of a bed in a general hospital, established at US $1047 [9], and the daily cost of a bed in a HACLES, estimated at US $373 [10]. Declared costs were adjusted to 2024 values considering increases in the cost of living.

### Data Analysis

Data curation included eliminating outliers in the variable of hospital stay duration to avoid distorting the used statistics. Specifically, outliers in the variable of hospital stay duration were identified using the IQR method. Observations falling below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR were considered outliers. These were removed to prevent distortion in the statistical analyses, ensuring a more accurate reflection of the central tendency and variability.

Descriptive statistics (mean and SD) were used to summarize data and identify patterns. Comparative analyses were conducted using both parametric and nonparametric tests, depending on the data distribution. The Student *t* test (2-tailed) was used to compare men and women for continuous variables, while the chi-square test was used for qualitative variables such as identifying patterns in the origin and destination of patients after discharge. The Cochran-Mantel-Haenszel test was conducted to analyze gender-adjusted differences in assistance rates. Nonparametric tests such as Kruskal-Wallis and Wilcoxon tests were used when assumptions for parametric tests were not met. Statistical significance was set at $P<.05$ (2-sided) for all tests.

### Ethical Considerations

The Ethics Committee of the Health Department of Alicante-Sant Joan approved the study (reference 24/044). The

Ethics Committee waived the requirement to obtain informed consent from all patients, as it was an epidemiological study that met the criteria of necessity, proportionality, and adequacy, offering comprehensive guarantees for the protection of personal data and respect for the privacy of the individuals involved, in accordance with Spanish Law 14/2007 on Biomedical Research.

## Results

Data from a total of 123 patients (73.2% of the expected total of 168) were recorded. Almost half of the sample were men (61/123, 49.6%). The mean age of the sample was 77.2 (SD 13.1) years. In 81 (65.9%) cases, the primary caregiver was a cohabiting family member (Table 1).

The average duration of the hospital stay was 83.3 (SD 80.5) days. Excluding outliers, the stay for men (55.2, SD 52.6 days) and women (53.1, SD 53.4 days) was similar (*P*=.21). At discharge, the number of daily medications administered to patients was 7.7 (SD 4) for men and 8.2 (SD 3.8) for women (*P*=.33). The reason for admission to the HACLES was similar between men (36/61, 59% palliative care; 14/61, 23% rehabilitation; 5/61, 8.2% convalescence; and 6/61, 9.8% long stay) and women (39/62, 62.9% palliative care; 14/62, 22.6% rehabilitation; 5/62, 11.3% convalescence; and 6/62, 3.2% long stay; *P*=.52).

**Table 1.** Sample description.

| Variable | Total (N=123) | Men (n=61, 49.6%) | Women (n=62, 50.4%) |
|---|---|---|---|
| **Age (years), mean (SD)** | 77.2 (13.1) | 79.2 (12.8) | 74 (11.8) |
| **BMI (kg/m$^2$), mean (SD)** | 23.9 (5.4) | 22.8 (4.8) | 25.4 (6.0) |
| **Primary caregiver (multiple options could be selected), n (%)** | | | |
| Institution | 2 (1.6) | 0 (0) | 2 (2.8) |
| Cohabiting caregiver | 81 (65.9) | 42 (58.3) | 39 (58.3) |
| Hired caregiver | 21 (17.1) | 10 (13.9) | 11 (15.3) |
| Self (loneliness) | 7 (5.7) | 3 (4.2) | 4 (5.6) |
| No identified caregiver | 9 (7.3) | 4 (5.6) | 5 (6.9) |
| Other (eg, supervision by descendants, support from a nongovernmental organization, etc) | 21 (17.1) | 10 (13.9) | 11 (15.3) |
| **Emergency contact, n (%)** | | | |
| Spouse | 44 (35.8) | 28 (45.9) | 16 (25.8) |
| Child | 54 (43.9) | 19 (31.1) | 35 (56.4) |
| Sibling | 13 (10.6) | 7 (11.5) | 6 (9.7) |
| Other family | 12 (9.8) | 7 (11.5) | 5 (8.1) |
| **Degree of dependency, n (%)** | | | |
| None | 57 (46.3) | 28 (45.9) | 29 (46.8) |
| Moderate | 0 (0) | 0 (0) | 0 (0) |
| Severe | 9 (7.3) | 4 (6.6) | 5 (8.1) |
| High dependency | 10 (8.1) | 3 (4.9) | 7 (11.3) |
| Under review | 47 (38.2) | 26 (42.6) | 21 (33.9) |
| **Receiving teleassistance, n (%)** | 6 (4.9) | 2 (3.3) | 4 (6.5) |
| **Reason for admission, n (%)** | | | |
| Palliative care | 75 (61) | 36 (59) | 39 (62.9) |
| Rehabilitation | 28 (22.8) | 14 (23) | 14 (22.6) |
| Convalescence | 12 (9.8) | 5 (8.2) | 7 (11.3) |
| Long-term stay | 8 (6.5) | 6 (9.8) | 2 (3.2) |
| **Risk of falling at admission, mean (SD)** | 73 (43.3) | 40 (26.2) | 33 (17.6) |
| **Number of chronic conditions at admission, mean (SD)** | 5.3 (4.1) | 4.8 (4.4) | 5.5 (4.2) |

In total, 70 patients died during their stay at the HACLES, while 54 were discharged and returned home (Table 2). The mortality rate was similar between men and women (men: 54.1%, 95% CI 47.8-71.5; women: 59.7%, 95% CI 42.2-66.0; *P*=.23). The degree of dependency at admission was higher in patients who

eventually died compared with those who did not (*P*=.008; Table 2).

**Table 2.** Factors related to mortality.

| Variable | Died during admission | | *P* value |
|---|---|---|---|
| | Yes (n=70) | No (n=53) | |
| **Scale, mean (SD)** | | | |
| Pfeiffer at admission | 4.3 (2.3) | 3.3 (3.2) | .11 |
| Barthel at admission | 27.4 (33.4) | 29.2 (33.2) | .75 |
| **Degree of dependency at admission[a], n (%)** | | | .008[b] |
| Moderate | 0 (0) | 0 (0) | |
| Severe | 4 (5.7) | 5 (9.4) | |
| High dependency | 4 (5.7) | 6 (11.3) | |
| None | 45 (64.3) | 12 (22.6) | |
| **Reason for admission, n (%)** | | | .52 |
| Palliative care | 39 (55.7) | 36 (67.9) | |
| Rehabilitation | 14 (20) | 14 (26.4) | |
| Convalescence | 7 (10) | 5 (9.4) | |
| Long-term stay | 2 (2.9) | 6 (11.3) | |

[a]47 patients were under review and were not included.

[b]The chi-square test does not include patients under review.

The average length of stay for patients who died at the HACLES was 35.4 (SD 80.1) days for women and 48.4 (SD 44.8) days for men (*P*=.40).

A total of 57 (46.3%) of the 123 patients admitted to the HACLES had a nonfamily caregiver or were institutionalized. Among those cared for by a cohabiting family member (81/123, 65.9%), the recognized degree of administrative dependency was higher (mean 1.3, SD 1.1 vs mean 1.9, SD 0.7; *P*=.001).

At discharge, a higher number of women than men were referred to another residential institution (11/25, 44% vs 7/28, 25%), although the difference was not statistically significant (*P*=.16). Most returned to their home (men: 21/28, 75%; women: 13/25, 52%; *P*=.11; Table 3). The length of stay was significantly larger for patients who returned home compared with patients discharged to other facilities (mean 89.8, SD 56.8 vs mean 33.1, SD 36.9 days; *P*<.001).

**Table 3.** Origin and destination of patients treated in the HACLES (Hospitals for Acute and Chronic Long-Term Extended Stay).

| Residence | At admission (N=123), n (%) | At discharge (N=123), n (%) | Men (n=61), n (%) | Women (n=62), n (%) |
|---|---|---|---|---|
| Home | 113 (91.9) | 34 (27.6) | 21 (34.4) | 13 (21) |
| Institution | 2 (1.6) | 18 (14.6) | 7 (11.5) | 11 (17.7) |
| Day center | 5 (4.1) | 1 (0.8) | 0 (0) | 1 (1.6) |
| Assisted living | 1 (0.8) | 0 (0) | 0 (0) | 0 (0) |
| Not recorded | 2 (1.6) | 0 (0) | 0 (0) | 0 (0) |
| Death | —[a] | 70 (56.9) | 33 (54.1) | 37 (59.7) |

[a]Not applicable.

Scores on the Barthel scale at discharge were 4 points higher than at admission, but this difference was not statistically significant (*P*=.46; Table 4). On the Pfeiffer scale, scores were higher at the beginning compared to discharge, with an initial score of 3.2 (SD 2.8) versus 2.5 (SD 2.8) at discharge (*P*=.03). At discharge, the Barthel score was the same for men and women (mean 36.4, SD 29.9 vs mean 30.2, SD 29.9; *P*=.45), while the Pfeiffer score for men was lower than for women (mean 1.7, SD 2.5 vs mean 3.9, SD 3.3; *P*=.01). The Gijón social-familial scale score was very similar for men and women at discharge (men: mean 7.6, SD 3.2; women: mean 8.3, SD 3.8; *P*=.27).

**Table 4.** Scores on the Barthel, Pfeiffer, and Gijón scales.

| Scale at admission | At admission, mean (SD) | At discharge, mean (SD) | Men at discharge, mean (SD) | Women at discharge, mean (SD) |
|---|---|---|---|---|
| Barthel (autonomy) | 29.7 (29.5) | 33.5 (26.1)[a] | 36.4 (29.9) | 30.2 (29.9) |
| Pfeiffer (cognitive function) | 3.2 (3.2) | 2.5 (3.0) | 1.7 (2.5) | 3.9 (3.3) |
| Gijón (socio-familial support) | —[b] | 8 (3.2) | 7.6 (2.9) | 8.3 (3.5) |

[a]*P*=.04.

[b]Not applicable.

Table 5 summarizes the estimated cost savings represented by the HACLES, both for the study sample and the extrapolation per 1000 admitted patients, for both men and women, considering the average stay of this study. On average, for every 1000 patients admitted to HACLES, the total hospital stay cost is reduced by US $42,614,846 (considering 61/123, 49.6% men and 62/123, 50.4% women). If the average length of stay is not corrected by eliminating outliers (average stay of 98.6 days for male patients and 69 days for female patients calculated during the study period), the cost differential for admissions in a general hospital compared with HACLES would be US $56,877.01.

**Table 5.** Cost savings from admissions in HACLES (Hospitals for Acute and Chronic Long-Term Extended Stay) compared with a general hospital[a].

| | Men | Women |
|---|---|---|
| Average stay (days), mean (SD) | 98.6 (119.9) | 69.0 (82.7) |
| Estimated savings per 1000 patients (US $ millions) | 46.3 | 39.8 |

[a]Average daily savings per bed of US $6280 compared with a general hospital.

## Discussion

### Principal Findings

The data from this study reflect that HACLES fulfill their function by offering an alternative admission option for patients requiring diversified and varying intensity care but for a longer duration, reducing the higher per-bed costs of general hospitals. The average age of patients admitted to a HACLES is around 77 years, with no significant differences between men and women. These patients have, on average, about 5 chronic conditions. In this sample, a significant proportion were admitted to the HACLES for palliative care, although almost a quarter were admitted for rehabilitation treatment in special conditions.

The average duration of stay in these centers is about 2 months, significantly different from the average of 7.5 days in general hospitals [11]. In this study, men had a 9.6-day longer stay than women, but this difference is not statistically significant due to the wide range of days of stay (including some stays up to 591 days).

HACLES could be responsible for saving over US $650 per patient per day of hospitalization, which means that both in terms of their specialization for the described patient profile and in economic terms, their existence is justified.

### Interpretation of Findings

Patients in long-term care facilities tend to be older and have a higher burden of chronic illnesses and comorbidities compared with the broader older adult population. Among the patients, 15 (15%) out of 100 had formal recognition of their level of dependency and were receiving state aid to manage their situation, with this being slightly more common among women than men. However, this does not imply that the patients admitted to HACLES had a less severe profile compared with other patients analyzed in different studies [12,13]; rather, it reflects the slow administrative process. Furthermore, it should be noted that almost 4 out of 10 patients admitted to HACLES are still awaiting evaluation of their degree of dependency due to significant delays in Spain's evaluation, registration, and subsidy assignment process, which can extend for over a year and a half.

Slightly more than half of the patients admitted died, consistent with or slightly lower than findings reported in similar studies [12]. This result contrasts with the lower mortality rates reported in studies conducted in general hospitals [14] and among patients with severe cognitive decline admitted to critical care units [15].

Most of those who died did not have recognized dependency, although a part (27/70, 38.2%) was pending resolution or under study. The reason for admission was not related to the mortality rate. Among those with formal administrative recognition of dependency, this situation does not account for differences in mortality, length of stay, or discharge destination. The mortality rate in this study was practically the same between men and women, as highlighted in other studies conducted in contexts more comparable to this one [16].

Scores on the Barthel scale indicated that both men and women generally showed severe dependency. In this sample, women presented greater cognitive deterioration than men admitted during the study period. Greater functional impairment in women, along with a similar mortality rate between men and women, have been reported in other studies [17], and our data confirm this trend. In this case, while women's social risk was moderate, men faced lower risk, which could explain their higher proportion of home returns upon discharge. However, it should be noted that the life expectancy for men (84.8 years) and women (88.3 years) in Spain [18] for this age group is significantly different.

The frequency with which women are referred to a residential institution upon discharge instead of returning home requires attention and new studies with larger samples. The trend in our data, combined with scores on autonomy, cognitive function, and social and family relationships, suggests that the difference may be due to women more frequently assuming the role of caregivers compared with men because of gender bias ingrained in our society. Other studies suggest that this is a phenomenon that extends beyond Spain [19]; thus, when planning alternative long-term care services, the different needs and preferences of men and women [19,20] should be taken into consideration.

The main reason for organizing care for this patient profile around a HACLES remains unchanged. Nearly 30 years after its implementation [21], its function is still necessary, and with the increase in chronic conditions, its capacity will likely be limited to respond to demand. The care needs required by older patients with multiple chronic conditions and limitations in their autonomy, along with the cost savings these centers have compared with general hospital admissions, corroborate the function of these centers, in line with what happens in other countries [22,23]. Furthermore, not only is the cost per bed lower but it is also important to consider, as some studies point out [24], that the number of medical tests and therapeutic interventions for this patient profile is higher when admitted to an acute care hospital. This increases risks for patients, affects their well-being, and raises the cost of care.

In the literature on long-term care hospitals and chronic patient management, the redistribution of spending across care settings has long been recognized usually as a decisive factor in cost savings [25,26]. The reinvestment of these savings should contribute to the expansion of services for older adults. Management models like the Chronic Care Model [27] have been adopted in various countries [28], including Spain [29], to provide a comprehensive approach to care. This aspect is crucial, as highlighted by the results of this study.

## Practical Implications

The health care teams in HACLES and their management teams could review their action protocols, promoting integrated actions with community resources to ensure a return home that addresses the social and personal needs of the people they serve, particularly in the case of women who show a greater tendency toward institutionalization after discharge. It should not be ruled out that while women commonly assume the role of informal caregivers for their male partners, men who take on this caregiving role face more limitations due to having less experience with domestic tasks. In this context, developing new approaches to involve relatives in long-term care is a significant challenge in many countries. Engaging family members effectively can improve patient outcomes and overall well-being, but it requires innovative strategies and resources to support both the patients and their caregivers [30].

Most patients were cared for by a cohabiting relative, typically the spouse or 1 or more children. Almost a quarter might present loneliness, which is in line with data analyzing the frequency of loneliness in the general population in Spain [31], despite almost half of this group receiving help from a non-cohabiting caregiver. This data suggest that one function HACLES could assume is to address, before discharge and in collaboration with community or social welfare resources, measures to limit the impact of loneliness upon return home. This would help address a growing social problem and reduce primary care costs [32].

The study findings have further implications for health care policies. First, since these facilities meet a significant practical need, particularly for older adult patients requiring prolonged hospital care at a much lower cost than general hospitals, health care planners should consider this approach as a viable alternative due to its efficiency in both public health systems, like Spain's, and in other mixed or predominantly private systems. Second, health care planners and managers of health care structures, such as primary and community care, should work together to address the challenges of integrating care after discharge from a long-term care hospital, as has already been suggested [33]. Home care should be considered to maintain positive outcomes for as long as possible after leaving the long-term care facility, thereby preventing readmissions.

## Future Research

Given the high mortality rate within a year after discharge [16], it should be analyzed in relation to the discharge destination of these patients, monitoring differences between men and women and considering the different options available to each group.

The impact of HACLES activities diminishes if adequate postdischarge care is not ensured. This requires offering integrated care [28] that addresses the needs and preferences of both men and women during this phase of their lives. Further research providing insights into the critical success factors of these interventions would facilitate more rapid implementation of effective solutions when patients are discharged from long-term care facilities.

## Strengths

This study is conducted in a field where the number of studies is limited, particularly in Spain. It describes the profile of patients treated in HACLES, analyzing gender differences. It identifies potential gender bias in home care provision. It also analyzes cost differences with general hospitals.

## Limitations

This study was conducted in a HACLES in one Autonomous Community in Spain. There may be long-term care centers organized to treat patients with other chronic conditions that do not fit the profile of patients treated in this center. Information records in clinical histories may sometimes be incomplete. When interpreting these data, the impact of the long stay on postadmission mortality should be considered [34].

## Conclusion

In summary, HACLES contribute to the sustainability of the health care system, and the reasons for their creation remain valid. The profile of needs of the patients admitted to these centers requires an integral approach, combining clinical, rehabilitative, psychological, and social aspects, not only during admission but also to set up appropriate mechanisms and resources according to needs, to offer integrated quality care to this generally more vulnerable group.

## Acknowledgments

This study was made possible thanks to the support of Hospital Padre Jofre. This manuscript was originally written in Spanish and translated into English by the authors with the assistance of GPT-4.0 (OpenAI) [36], a large language model, to ensure readability in English. The authors reviewed and edited the content as needed and take full responsibility for the final version of the publication.

This study was funded by the Consellería de Innovación, Universidades, Ciencia y Sociedad Digital, Generalitat Valenciana, PROMETEO 2021/061. During the study period, JJM received a grant for intensifying research activity from the Instituto de Salud Carlos III, reference INT22/00012.

## Data Availability

The datasets generated during and/or analyzed during this study are available in the Open Science Framework repository [35].

## Authors' Contributions

JJM and CC conceptualized the study. DG curated the data and conducted the data analysis. MMB, AIC, MR, and RM carried out the field study and participated, along with the other authors, in the interpretation of the results. JJM and DG wrote the original draft. All authors reviewed, edited, and approved the final version.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[DOCX File , 33 KB - ijmr_v13i1e64248_app1.docx ]

## References

1. Doty P. Long-term care in international perspective. Health Care Financ Rev 1988;Spec No(Suppl):145-155 [FREE Full text] [Medline: 10312969]
2. Perracini MR, Arias-Casais N, Thiyagarajan JA, Rapson C, Isaac V, Ullah S, et al. A recommended package of long-term care services to promote healthy ageing based on a WHO global expert consensus study. J Am Med Dir Assoc 2022;23(2):297-303.e14. [doi: 10.1016/j.jamda.2021.12.019] [Medline: 34973958]
3. Kahn JM, Barnato AE, Lave JR, Pike F, Weissfeld LA, Le TQ, et al. A comparison of free-standing versus co-located long-term acute care hospitals. PLoS One 2015;10(10):e0139742. [doi: 10.1371/journal.pone.0139742] [Medline: 26440102]
4. Avilés MJ, Cuevas MD, Zafra E. Strategy for the care of chronic patients in the Valencian Community. Estrategia para la atención a pacientes crónicos en la Comunitat Valenciana. Consellería Sanitat. 2014. URL: https://www.san.gva.es/es/web/assistencia-sanitaria/estrategia-para-la-atencion-a-pacientes-cronicos-en-la-comunitat-valenciana [accessed 2024-10-30]
5. Zimmerman S, Cesari M, Gaugler JE, Gleckman H, Grabowski DC, Katz PR, et al. The inevitability of reimagining long-term care. J Am Med Dir Assoc 2022;23(2):187-189. [doi: 10.1016/j.jamda.2021.12.032] [Medline: 35123697]
6. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. Md State Med J 1965;14:61-65 [FREE Full text]
7. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975;23(10):433-441. [doi: 10.1111/j.1532-5415.1975.tb00927.x] [Medline: 1159263]
8. Díaz-Palacios JM, Alarcón T. Validation of a five-item version of the Geriatric Depression Scale in Spanish population. Validación de una versión modificada de la escala socio-familiar de Gijón (SFES). Revista Española de Geriatría y Gerontología 1994;29:323-328. [doi: 10.1016/s0211-139x(01)74736-1]
9. Indexes by autonomous communities: general and ECOICOP groups. Índices nacionales: general y de grupos ECOICOP. Instituto Nacional de Estadística. 2024. URL: https://www.ine.es/jaxiT3/Tabla.htm?t=50913&L=0 [accessed 2024-10-30]
10. Healthcare in Spain: a guide to the Spanish healthcare system. Wise. URL: https://wise.com/us/blog/healthcare-system-in-spain [accessed 2024-10-30]
11. OECD. Health at A Glance 2023: OECD Indicators. Paris, France: OECD Publishing; 2023.
12. Silva TJA, Jerussalmy CS, Farfel JM, Curiati JAE, Jacob-Filho W. Predictors of in-hospital mortality among older patients. Clinics (Sao Paulo) 2009;64(7):613-618. [doi: 10.1590/S1807-59322009000700002] [Medline: 19606235]
13. Dalrymple U, Makwana A, Pring A, Bowtell N, Verne J. Older people's hospital admissions in the last year of life. GOV.UK. URL: https://www.gov.uk/government/publications/older-peoples-hospital-admissions-in-the-last-year-of-life/older-peoples-hospital-admissions-in-the-last-year-of-life [accessed 2024-10-30]

14. Vetrano DL, Landi F, de Buyser SL, Carfì A, Zuccalà G, Petrovic M, et al. Predictors of length of hospital stay among older adults admitted to acute care wards: a multicentre observational study. Eur J Intern Med 2014;25(1):56-62. [doi: 10.1016/j.ejim.2013.08.709] [Medline: 24054859]
15. Li HC, Chen CCH, Yeh TYC, Liao SC, Hsu AS, Wei YC, et al. Predicting hospital mortality and length of stay: a prospective cohort study comparing the intensive care delirium screening checklist versus confusion assessment method for the intensive care unit. Aust Crit Care 2023;36(3):378-384. [doi: 10.1016/j.aucc.2022.01.010] [Medline: 35272910]
16. Tanuseputro P, Chalifoux M, Bennett C, Gruneir A, Bronskill SE, Walker P, et al. Hospitalization and mortality rates in long-term care facilities: does for-profit status matter? J Am Med Dir Assoc 2015;16(10):874-883. [doi: 10.1016/j.jamda.2015.06.004] [Medline: 26433862]
17. Almagro P, Ponce A, Komal S, de la Asunción Villaverde M, Castrillo C, Grau G, et al. Multimorbidity gender patterns in hospitalized elderly patients. PLoS One 2020;15(1):e0227252. [doi: 10.1371/journal.pone.0227252] [Medline: 31990911]
18. Mortality tables. Tasas de mortalidad. Instituto Nacional de Estadística. URL: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177004&menu=ultiDatos&idp=1254735573002 [accessed 2024-10-30]
19. Mudrazija S, Thomeer MB, Angel JL. Gender differences in institutional long-term care transitions. Womens Health Issues 2015;25(5):441-449. [doi: 10.1016/j.whi.2015.04.010] [Medline: 26123639]
20. Carvalho N, Meylan L, Blanco JM, Fustinoni S, Abolhassani N, Santos-Eggimann B. Gender differences regarding opinions on long-term care arrangements: a study of community-dwelling older adults. Arch Gerontol Geriatr 2019;83:195-203 [FREE Full text] [doi: 10.1016/j.archger.2019.03.018] [Medline: 31082564]
21. Rentería RB. The challenge of chronicity in Spain: better transform rather than ration. El reto de la cronicidad en España: mejor transformar que racionar. Gac Sanit 2015;29(5):323-325 [FREE Full text]
22. Einav L, Finkelstein A, Mahoney N. Long-term care hospitals: a case study in waste. The Review of Economics and Statistics 2023;105:745-765. [doi: 10.1162/rest_a_01092]
23. Crawford R, Stoye G, Zaranko B. Long-term care spending and hospital use among the older population in England. J Health Econ 2021;78:102477. [doi: 10.1016/j.jhealeco.2021.102477] [Medline: 34153887]
24. Friedman R, Kalant N. Comparison of long-term care in an acute care institution and in a long-term care institution. CMAJ 1998;159(9):1107-1113. [Medline: 9835878]
25. Kaye HS, Harrington C, LaPlante MP. Long-term care: who gets it, who provides it, who pays, and how much? Health Aff (Millwood) 2010;29(1):11-21. [doi: 10.1377/hlthaff.2009.0535] [Medline: 20048355]
26. Balqis-Ali NZ, Jawahir S, Chan YM, Lim AWY, Azlan UW, Shaffie SSM, et al. The impact of long-term care interventions on healthcare utilisation among older persons: a scoping review of reviews. BMC Geriatr 2024;24(1):484. [doi: 10.1186/s12877-024-05097-9] [Medline: 38831269]
27. Wagner EH, Davis C, Schaefer J, Von Korff M, Austin B. A survey of leading chronic disease management programs: are they consistent with the literature? Manag Care Q 1999;7(3):56-66. [Medline: 10620960]
28. Forbes LJ, Marchand C, Doran T, Peckham S. The role of the quality and outcomes framework in the care of long-term conditions: a systematic review. Br J Gen Pract 2017;67(664):e775-e784. [doi: 10.3399/bjgp17X693077] [Medline: 28947621]
29. Mira JJ, Nuño-Solinís R, Guilabert-Mora M, Solas-Gaspar O, Fernández-Cano P, González-Mestre MA, et al. Development and validation of an instrument for assessing patient experience of chronic illness care. Int J Integr Care 2016;16(3):13. [doi: 10.5334/ijic.2443] [Medline: 28435422]
30. Gaugler JE, Mitchell LL. Reimagining family involvement in residential long-term care. J Am Med Dir Assoc 2022;23(2):235-240. [doi: 10.1016/j.jamda.2021.12.022] [Medline: 34973167]
31. Rodríguez Pérez M, Díaz-Olalla JM, Pedrero Pérez EJ, Sanz Cuesta MR. Informe monográfico: sentimiento de soledad en la ciudad de Madrid. In: Opioid Drug Use in the City of Madrid: Associated Health and Sociodemographic Factors. Estudio de Salud de la Ciudad de 2018 Madrid. Madrid, Spain: Madrid Salud, Ayuntamiento de Madrid; 2020:429-503.
32. Mira JJ, Torres D, Gil V, Carratalá C. Loneliness impact on healthcare utilization in primary care: a retrospective study. J Healthc Qual Res 2024;39(4):224-232. [doi: 10.1016/j.jhqr.2024.04.001] [Medline: 38670900]
33. Grudniewicz A, Gray CS, Boeckxstaens P, de Maeseneer J, Mold J. Operationalizing the Chronic Care Model with goal-oriented care. Patient 2023;16(6):569-578. [doi: 10.1007/s40271-023-00645-8] [Medline: 37642918]
34. López Pardo P, Socorro García A, Baztán Cortés JJ. Influence of length of hospital stay on mortality after discharge in older patients with acute medical diseases. Gac Sanit 2016;30(5):375-378. [doi: 10.1016/j.gaceta.2016.04.008] [Medline: 27266515]
35. García-Torres D. Patient profile and cost savings of long-term care in a Spanish hospital. Open Science Framework. 2024 Nov 13. URL: https://osf.io/2ugv4/?view_only=58a416002d574f2a8736e6259dac03db [accessed 2024-11-13]
36. GPT-4. OpenAI. URL: https://platform.openai.com/docs/models/gpt-4 [accessed 2024-11-12]


## Abbreviations

**HACLES:** Hospitals for Acute and Chronic Long-Term Extended Stay
**ICD-10:** International Classification of Diseases, 10th revision

**OR:** odds ratio
**STROBE:** Strengthening the Reporting of Observational Studies in Epidemiology

*Edited by G Eysenbach, T de Azevedo Cardoso; submitted 12.07.24; peer-reviewed by U Khan; comments to author 14.08.24; revised version received 15.08.24; accepted 28.10.24; published 19.11.24.*

*Please cite as:*
*Mira JJ, García-Torres D, Bonell-Guerrero MDM, Cáceres-Sevilla AI, Ramirez-Sanz M, Martínez-Lleo R, Carratalá C*
*Patient Profile and Cost Savings of Long-Term Care in a Spanish Hospital: Retrospective Observational Study*
*Interact J Med Res 2024;13:e64248*
*URL: https://www.i-jmr.org/2024/1/e64248*
*doi:10.2196/64248*
*PMID:*

Original Paper

# Splenectomy as a Risk Factor for Graft Rejection Following Endothelial Transplantation: Retrospective Study

Paola Kammrath Betancor[1], MD; Daniel Böhringer[1], Prof Dr; Philip Maier[1], Prof Dr; Thabo Lapp[1,2], Prof Dr; Thomas Reinhard[1], Prof Dr

[1]Eye Center, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[2]Ophtha-Lab, Department of Ophthalmology, St. Franziskus Hospital Muenster, Muenster, Germany

**Corresponding Author:**
Paola Kammrath Betancor, MD
Eye Center, Medical Center, Faculty of Medicine, University of Freiburg
Killianstrasse 5
Freiburg, 79106
Germany
Phone: 49 761 270 40010
Fax: 49 761 270 40630
Email: paola.kammrath.betancor@uniklinik-freiburg.de

## Abstract

**Background:** Anterior chamber–associated immune deviation (ACAID) is an active immunotolerance mechanism, which is induced by placing antigen into the anterior eye chamber as long as a major surgical trauma is avoided. For this reason, ACAID may be a major contributor to the favorable immunologic outcomes in Descemet membrane endothelial keratoplasty (DMEK). Rodent models have demonstrated the importance of a functional spleen for the development of an ACAID.

**Objective:** This study aimed to investigate whether splenectomy leads to increased rejection rates after DMEK in humans.

**Methods:** A retrospective evaluation was conducted on the course following DMEK at the Eye Center, Medical Center, University of Freiburg, for patients with a self-reported history of splenectomy compared to patients without this condition. Potential study patients were contacted by mail. A questionnaire to self-report splenectomy and the time thereof was sent out. The medical records of all consenting patients at the Eye Center were reviewed for graft survival and immune reactions.

**Results:** We asked 1818 patients after DMEK to report their history of splenectomy. A total of 1340 patients responded and were included in the study. Of these 1340 patients, 16 (1.2%) reported a history of splenectomy (ie, 26 DMEKs, with 10 patients being transplanted in both eyes and 6 patients being transplanted in 1 eye; median age at surgery 73.7, range 66.7-76.1 y). The remaining patients (1324 patients, ie, 1941 eyes) served as controls, with 1941 DMEKs (median age at surgery 71.5, range 64.1-77.2 y). Five (19%) out of the 26 eyes from the splenectomy group required a second transplant due to dislocation (n=2.8%), failure (n=2.8%), and rejection (n=1.4%). Kaplan-Meier analysis revealed no relevant difference compared with controls.

**Conclusions:** Our results suggest that splenectomy has no major effect on the outcome following DMEK. Subsequent, ACAID may not be the main reason for the favorable immunological outcomes in DMEK, or the camero-splenic axis may be subordinate in humans. However, we only included 16 patients who underwent splenectomy, so it might be possible that we missed a minor effect.

***(Interact J Med Res 2024;13:e50106)*** doi:10.2196/50106

**KEYWORDS**

anterior chamber–associated immune deviation; ACAID; Descemet membrane endothelial keratoplasty; DMEK; splenectomy

## Introduction

Nowadays, corneal transplantation is one of the most common and successful forms of tissue transplantation worldwide, and in the vast majority of cases, it occurs without a human leukocyte antigen match [1]. Since not only can penetrating keratoplasty be performed but lamellar surgery techniques (such as Descemet membrane endothelial keratoplasty [DMEK]) are also available, the proportion of penetrating keratoplasties is decreasing while the number of DMEKs performed is increasing [2]. Approximately 2% to 7% of normal-risk patients after a DMEK and about 18% after a penetrating keratoplasty

experience immunologic rejection [3-5]. At present, it remains unclear why there is less rejection in DMEK transplantation; consequently, a hypothesis arose—that the anterior chamber–associated immune deviation (ACAID) phenomenon may contribute significantly to the favorable immunologic outcomes in DMEK. As of today, there are practically no clinical data on this topic.

The fact that corneal transplantation can be effective was demonstrated by various surgeons and ophthalmologists in animal models as early as 1818 (on rabbits) [6]. The first successful corneal transplantation in humans was performed by Austrian surgeon Eduard Zirm [7] in 1906. In comparison, the first successful kidney transplantation was performed almost 50 years later, in 1953 [8]. Corneal transplants were much less likely to be rejected compared to organ transplants at very early stages after transplantation. The reason why the corneal tissue was not rejected directly led to many hypotheses about the immunology of the eye that allows this [9]. Considering these questions, clinicians as well as researchers began to investigate ACAID [10,11]. The immune system is able to prevent immune reactions against foreign antigens within the eye. In the rodent model, it has been found that this principle works only if the antigens are injected atraumatically into the eye [12]. The removal of the spleen prevents the development of ACAID in the rodent model, resulting in an increased rejection rate after corneal transplantation [13]. This is currently explained by the so-called camero-splenic axis. Antigen-presenting cells most likely derived from the iris and ciliary body take up antigens placed in the anterior chamber and migrate via the trabecular meshwork and collector veins through the blood into the spleen. Within the spleen, these antigen-presenting cells induce the differentiation of antigen-specific regulatory T cells, forming the cellular arm of ACAID [14]. However, there is currently no proof of this theory; human data on the presence of ACAID or the camero-splenic axis are lacking.

The primary objective of this retrospective study was to compare outcomes between human patients with and without splenectomy, specifically graft rejection and failure, following DMEK surgery. For this purpose, all reachable patients who underwent DMEK at the Eye Center, Medical Center, University of Freiburg, during the last 12 years (2734 eyes in total) were sent a questionnaire to self-report whether they underwent splenectomy and the time thereof. We herein investigate the potential role of ACAID in the excellent immunological prognosis of DMEK. This could provide initial evidence on the contribution of ACAID to the success rates of DMEK.

## Methods

### Study Design

This was a retrospective study evaluating the course following DMEK at the Eye Center, Medical Center, University of Freiburg, for patients with a self-reported history of splenectomy compared to patients without this condition.

### Data Collection

A total of 2734 DMEKs were performed between 2010 and 2022 at the Eye Centre, Medical Center, University of Freiburg. Patients who underwent DMEK during this period were eligible for inclusion. We were able to contact 1818 patients from this pool. They were sent a cover letter by mail, including patient information about the study, a consent form, and a questionnaire to self-report splenectomy and the time thereof. This questionnaire formed the basis of consent in our study. Patients who did not respond to the questionnaire or did not wish to participate in the study were excluded from the analysis.

The exposure of interest was a self-reported history of splenectomy, which was assessed through the questionnaire sent to patients. Data collected through the questionnaire included self-reported history of splenectomy and the time thereof.

The questionnaire was a simple 1-question survey asking patients to self-report whether they had undergone splenectomy and when. The questionnaire was developed by the research team and reviewed and approved by the ethics committee to ensure clarity and comprehensibility.

After submitting the questionnaire, patients cannot withdraw their participation in the research project, as we irreversibly anonymize their identity and destroy their questionnaire in accordance with data protection regulations once their information have been transferred for the statistical analysis. Therefore, individual study participants cannot be identified when publishing the study results.

Data collected from electronic health records included demographic information, indication for DMEK, and postoperative outcomes such as graft rejection and failure.

The medical records of all consenting patients were reviewed for graft survival and immune reactions. All patients are scheduled postoperatively for follow-up in our clinic, so it is highly likely that we have captured almost all immune reactions. These data were linked to the questionnaire responses to compare patients who underwent splenectomy to the remainder who served as controls.

DMEK surgery was executed per standards as previously described [3]. Briefly, a trephine was used to punch the grafts, which were stained with trypan blue 0.6 mg/mL. Subsequent to descemetorhexis and insertion of the graft into the anterior chamber, the graft was unfolded and centered. Next, complete air filling of the anterior chamber connected the graft to the posterior stroma. Postoperatively, patients were requested to remain supine for 3 days. Postoperative local therapy consisted of topical dexamethasone 5 times daily and tapered over 5 months to once daily, which was recommended for up to 24 months or longer. There was no evidence of any difference in postoperative medical aftercare between patients who under splenectomy and the controls.

Patients with postoperative epithelial defects were initially treated with dexpanthenol and ofloxacin ointment alternately every hour until epithelial closure, before the aforementioned treatment was started.

The main outcomes were graft rejection and failure.

Graft rejection was determined by reviewing medical records for signs of rejection, such as newly appearing endothelial

precipitates on the graft [5]. Other clinical signs of graft rejection may include cells in the anterior chamber or otherwise unexplained corneal edema [15].

Corneal grafts that were not adherent and/or did not result in corneal transparency were classified as graft failures.

### Ethical Considerations

The study protocol was approved by the ethics committee of the Albert Ludwigs University of Freiburg (21-1472). Informed consent was obtained from all participants, and they had the ability to opt out of the study. Data were anonymized, and protective measures were in place to safeguard participant information. No compensation was provided to participants in this study.

### Data Presentation and Statistical Analysis

Descriptive statistics summarized baseline characteristics between the splenectomy and control groups. Median and IQR were reported for continuous variables, and percentages were reported for categorical variables. Differences were assessed using ANOVA for continuous data and Pearson chi-square tests for categorical outcomes.

Time-to-event analysis was performed using Kaplan-Meier curves depicting the risk of immune reactions, graft failure (as operationalized by regrafting), and rejection-free graft survival (a combination of the 2 aforementioned end points) between the splenectomy and control groups. Additionally, we also conducted another analysis by excluding early regrafting (during the first few postoperative days) due to technical failures. Groups were compared using the log-rank test. Multivariable Cox proportional hazard models were constructed to determine the independent association of splenectomy on outcomes after adjusting for potential confounding factors such as recipient age, sex, baseline diagnosis, and triple surgery (ie, cataract surgery in combination with DMEK). Hazard ratios with 95 CIs were reported.

Two-sided *P* values <.05 were considered statistically significant for all analyses, which were performed with R software (version 4.1.3; R Foundation for Statistical Computing) [16].

## Results

Of the 1818 patients contacted, 1311 (72.1%) responded by returning the questionnaire. Additionally, 11 (0.6%) patients responded by email, and 18 (1%) responded via telephone. This yielded a response rate of 73.7% (1340/1818). Among the 1340 respondents, 16 (1.2%) reported a history of splenectomy. Only 5 other patients reported back and did not want to be included in the study. The family members of 27 patients contacted us to inform us that the patient had died and that no information about splenectomy was available; thus, these patients were not included in the study.

Of the 1324 patients who did not undergo splenectomy, we performed DMEKs on both eyes among 617 patients, so a total of 1941 eyes could be included in the control group.

Of the patients who underwent splenectomy, some had received DMEKs on both eyes at our hospital, yielding a total of 26 DMEK procedures. Of the 16 patients who underwent splenectomy, 11 (69%) were able to recall the time of the surgery. The remaining 5 (31%) patients could not give information about the time of surgery. Only 2 patients (each having had a DMEK on only 1 eye) underwent the splenectomy after the DMEK (see Multimedia Appendix 1). Median age at transplantation was similar between the groups. However, there was a lower proportion of female patients (9/26, 35% vs 1129/1941, 58.2%) and a higher rate of triple surgery (20/26, 77% vs 1139/1941, 58.7%) in the splenectomy group.

After reviewing the medical records, we found that of the 26 DMEKs (16 patients) in the splenectomy group, 5 (19%) eyes required regrafting. Two of these repeated DMEKs were performed for technical reasons because of incomplete graft attachment. Only 1 patient presented the defined signs of endothelial graft rejection, and the remaining 2 patients showed unspecific late endothelial graft failure (see Table 1 and Multimedia Appendix 2).

**Table 1.** Summative evaluation of the Descemet membrane endothelial keratoplasties (DMEKs) of the 16 patients who underwent splenectomy.

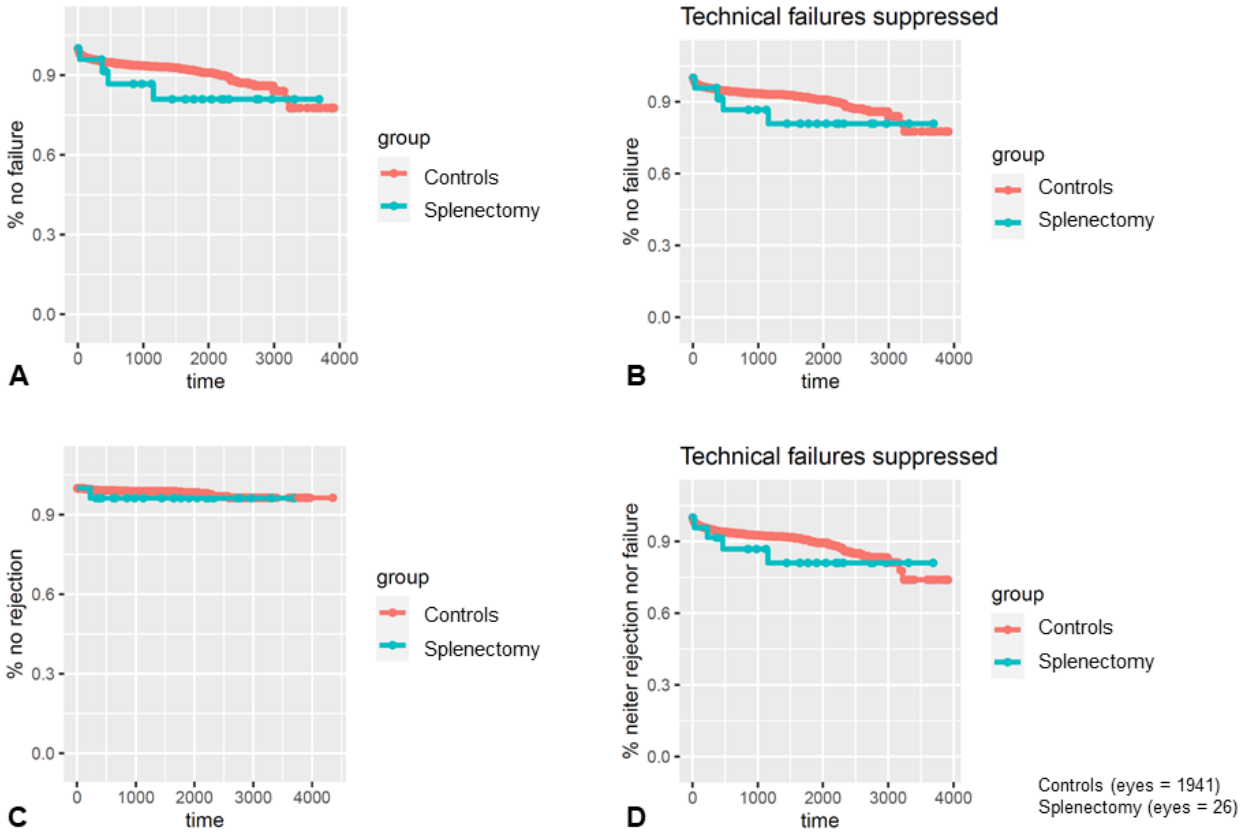
| Variable | Value |
|---|---|
| **Patients who underwent splenectomy, n** | 16 |
| DMEKs, n | 26 |
| **Repeat DMEKs (n=26 eyes), n (%)** | 5 (19) |
| Graft dislocation | 2 (8) |
| Graft failure | 2 (8) |
| Graft rejection | 1 (4) |

In the control group, graft failure occurred in 153 (7.9%) out of 1941 eyes, and rejections were observed in 26 (1.3%) eyes. We additionally compared the indication for DMEK between the patients who underwent splenectomy and the controls (see Multimedia Appendix 2). Of the 26 DMEKs in patients who underwent splenectomy, 24 (92%) had Fuchs endothelial dystrophy. Two patients had previously received penetrating keratoplasty and required a DMEK for graft failure. Precisely, these 2 cases required a second DMEK, both due to endothelial graft failure.

Kaplan-Meier analysis (Figure 1) and multivariable Cox proportional hazards models adjusting for age, sex, baseline

diagnosis, and triple surgery revealed no statistically significant differences in the risk of regrafting, failure, rejection, or combined end points between groups (all $P>.05$).

**Figure 1.** Survival analyses. (A) The percentage of patients who did not need a repeat keratoplasty shown over time in days. The patients who underwent splenectomy are presented in blue and the control group is shown in red. The splenectomy group includes the 2 cases of graft dislocation. (B) Without these technical failures, a slight trend of patients who underwent splenectomy toward the need of a second keratoplasty may be seen. The same plots are shown for (C) the rejection-free time and (D) without the technical failures. Included in these lower 2 curves is 1 rejection case that did not require repeat keratoplasty. Again, one could assume a trend here.




Survival analyses comparing graft rejection rates, the need for repeat keratoplasties, and the combined results between groups are shown in Figure 1. The control group is considerably larger than the splenectomy group for obvious reasons. When counting the early repeat DMEKs from graft dislocation, the splenectomy group showed a trend toward a slightly worse performance compared to the controls (see Figure 1A). However, this is abrogated when counting only events that could theoretically be caused by immune reactions (see Figure 1B). In the comparison of the DMEK grafts without rejection reaction, almost no difference between the groups is discernable (see Figure 1C). Ignoring technical failures when regarding the combined results (see Figure 1D), a slight trend may be inferred toward the splenectomy group performing worse than the controls.

Specifically, the adjusted hazard ratio for the risk of graft failure comparing patients who underwent splenectomy to controls was 1.91 (95% CI 0.25-14.31; $P$=.53). Similar nonsignificant findings were observed for rejection and combined outcome models.

## Discussion

### Principal Findings

Our study was designed to investigate the effects of splenectomy on rejection after DMEK. Based on the data presented here, the spleen does not appear to have a major influence on the survival or rejection of DMEK grafts, so the significance of ACAID for human DMEK may be subordinate.

Animal models theorize that the spleen is essential in the induction of ACAID [13,14]; evidence to support this theory does not exist from clinical studies. This study is the first large-scale study to examine this in a retrospective clinical setting. We found no evidence to support these hypotheses.

It is still unclear whether ACAID and immune privilege for corneal transplantation established in animal models also applies to human keratoplasty and more specifically to DMEK. The low risk of allograft rejection after corneal endothelial transplantation is thought to be due to the ACAID phenomenon after the introduction of antigens into the anterior chamber [17]. As early as 1966, Streilein et al [18] demonstrated that external corneal procedures, such as keratoplasty, corneal cauterization, and corneal sutures, lead to inflammation that prevents the induction of ACAID. Yamada et al [19] examined the allogeneic

response in the anterior chamber after the transplantation of corneal endothelial cells in a mouse model. Both intracameral injection of splenocytes and corneal endothelial cells induced ACAID with the suppression of the delayed hypersensitivity reaction. However, this could not be detected in inflamed eyes by cryoinjury, so the loss of the delayed hypersensitivity reaction does not seem to be regulated by ACAID.

Not all animal studies showed an effect of splenectomy on corneal graft survival. Bourne et al [20] performed corneal transplants in 19 rabbits each with and without splenectomy and found no significant difference in graft survival. In 2017, Vendomèle et al [11] summarized the evidence on ACAID and noted that several factors raised questions about the reliability and validity of studies using knockout mouse models. In particular, physiological relevance and transferability to humans must be considered critically.

To our knowledge, our study is the first with such a high number of patients who underwent both splenectomy and DMEK. Hos et al [21] followed a single case of a patient who underwent splenectomy for 4 years after DMEK. During this time, they noted no corneal graft rejection and assumed that the spleen was not necessary for graft acceptance.

Of our 26 DMEK transplants, 1 case developed graft rejection and 2 cases developed graft failure. In relation to the control group, this may suggest that splenectomy possibly does lead to a slight tendency of poorer graft acceptance. However, it has to be considered that the 2 graft failures are part of the high-risk group, because the indication for DMEK was the failure of a previously performed penetrating keratoplasty. Thus, from the data presented here, splenectomy does not appear to have a clear effect on the immunologic response to corneal transplantation. Furthermore, it should be noted that of all participating patients, only 16 patients who underwent splenectomy were included; thus, it is possible that we missed finding an influence on graft rejection after DMEK. The validity and reliability of the basic questionnaire was not formally tested. This is a limitation given the self-reported nature of the splenectomy history but is mitigated by the minimalistic (only 1 question) nature of this document. Additionally, we can conclude that undergoing splenectomy after DMEK does not have any influence either, since 2 of our patients showed such a course.

## Conclusion

As described previously, there is controversy about the relevance of the spleen in animal models investigating corneal graft rejection. The importance of the spleen in humans has not yet been investigated. Our results suggest that splenectomy does not substantially impact DMEK outcomes after accounting for potential confounding factors. However, given the limited sample size of patients who underwent splenectomy, a clinically meaningful effect cannot be definitively excluded. Nonetheless, ACAID may either not fully explain favorable immunologic outcomes following DMEK based on current evidence or the camero-splenic axis may be subordinate in humans.

## Acknowledgments

We acknowledge support by the Open Access Publication Fund of the University of Freiburg.

## Authors' Contributions

DB and TR contributed to conceptualization and design. PKB and DB contributed to data curation. DB, PKB, and TR contributed to methodology. DB contributed to software. PKB and DB contributed to formal analysis. DB and TR contributed to supervision. PKB, DB, and TR contributed to writing—original draft. PKB, DB, PM, TL, and TR contributed to writing—review and editing.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Time course of the Descemet membrane endothelial keratoplasties (DMEKs) of the 16 patients who underwent splenectomy, including the indication for the DMEK.
[DOCX File , 18 KB - ijmr_v13i1e50106_app1.docx ]

Multimedia Appendix 2
Summarized data of the control and splenectomy groups.
[DOCX File , 16 KB - ijmr_v13i1e50106_app2.docx ]

## References

1. van Essen TH, Roelen DL, Williams KA, Jager MJ. Matching for human leukocyte antigens (HLA) in corneal transplantation - to do or not to do. Prog Retin Eye Res 2015 May;46:84-110. [doi: 10.1016/j.preteyeres.2015.01.001] [Medline: 25601193]
2. Flockerzi E, Maier P, Böhringer D, Reinshagen H, Kruse F, Cursiefen C, et al. Trends in corneal transplantation from 2001 to 2016 in Germany: a report of the DOG-Section Cornea and its keratoplasty registry. Am J Ophthalmol 2018 Apr;188:91-98 [FREE Full text] [doi: 10.1016/j.ajo.2018.01.018] [Medline: 29410297]

3. Heinzelmann S, Böhringer D, Eberwein P, Reinhard T, Maier P. Outcomes of Descemet membrane endothelial keratoplasty, Descemet stripping automated endothelial keratoplasty and penetrating keratoplasty from a single centre study. Graefes Arch Clin Exp Ophthalmol 2016 Mar;254(3):515-522. [doi: 10.1007/s00417-015-3248-z] [Medline: 26743748]
4. Anshu A, Price MO, Price FW. Risk of corneal transplant rejection significantly reduced with Descemet's membrane endothelial keratoplasty. Ophthalmology 2012 Mar;119(3):536-540. [doi: 10.1016/j.ophtha.2011.09.019] [Medline: 22218143]
5. Hos D, Tuac O, Schaub F, Stanzel TP, Schrittenlocher S, Hellmich M, et al. Incidence and clinical course of immune reactions after Descemet membrane endothelial keratoplasty: retrospective analysis of 1000 consecutive eyes. Ophthalmology 2017 Apr;124(4):512-518. [doi: 10.1016/j.ophtha.2016.12.017] [Medline: 28094043]
6. Payne JW. New directions in eye banking. Trans Am Ophthalmol Soc 1980;78:983-1026 [FREE Full text] [Medline: 7020217]
7. Zirm EK. Eine erfolgreiche totale Keratoplastik (a successful total keratoplasty). 1906. Refract Corneal Surg 1989;5(4):258-261. [Medline: 2488816]
8. Hatzinger M, Stastny M, Grützmacher P, Sohn M. The history of kidney transplantation [Article in German]. Urologe A 2016 Oct;55(10):1353-1359. [doi: 10.1007/s00120-016-0205-3] [Medline: 27518791]
9. Medawar PB. Immunity to homologous grafted skin; the fate of skin homografts transplanted to the brain, to subcutaneous tissue, and to the anterior chamber of the eye. Br J Exp Pathol 1948 Feb;29(1):58-69 [FREE Full text] [Medline: 18865105]
10. Streilein JW. Anterior chamber associated immune deviation: the privilege of immunity in the eye. Surv Ophthalmol 1990;35(1):67-73 [FREE Full text] [doi: 10.1016/0039-6257(90)90048-z] [Medline: 2204129]
11. Vendomèle J, Khebizi Q, Fisson S. Cellular and molecular mechanisms of anterior chamber-associated immune deviation (ACAID): what we have learned from knockout mice. Front Immunol 2017 Nov 30;8:1686 [FREE Full text] [doi: 10.3389/fimmu.2017.01686] [Medline: 29250068]
12. Wilbanks GA, Streilein JW. Characterization of suppressor cells in anterior chamber-associated immune deviation (ACAID) induced by soluble antigen. evidence of two functionally and phenotypically distinct T-suppressor cell populations. Immunology 1990 Nov;71(3):383-389. [Medline: 1702748]
13. Niederkorn JY, Mellon J. Anterior chamber-associated immune deviation promotes corneal allograft survival. Invest Ophthalmol Vis Sci 1996 Dec;37(13):2700-2707. [Medline: 8977485]
14. Streilein JW. Ocular immune privilege: therapeutic opportunities from an experiment of nature. Nat Rev Immunol 2003 Nov;3(11):879-889. [doi: 10.1038/nri1224] [Medline: 14668804]
15. Hos D, Matthaei M, Bock F, Maruyama K, Notara M, Clahsen T, et al. Immune reactions after modern lamellar (DALK, DSAEK, DMEK) versus conventional penetrating corneal transplantation. Prog Retin Eye Res 2019 Nov;73:100768 [FREE Full text] [doi: 10.1016/j.preteyeres.2019.07.001] [Medline: 31279005]
16. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2016-08-07]
17. Niederkorn JY. The induction of anterior chamber-associated immune deviation. Chem Immunol Allergy 2007;92:27-35. [doi: 10.1159/000099251] [Medline: 17264480]
18. Streilein JW, Bradley D, Sano Y, Sonoda Y. Immunosuppressive properties of tissues obtained from eyes with experimentally manipulated corneas. Invest Ophthalmol Vis Sci 1996 Feb;37(2):413-424. [Medline: 8603847]
19. Yamada J, Ueno M, Toda M, Shinomiya K, Sotozono C, Kinoshita S, et al. Allogeneic sensitization and tolerance induction after corneal endothelial cell transplantation in mice. Invest Ophthalmol Vis Sci 2016 Sep 01;57(11):4572-4580. [doi: 10.1167/iovs.15-19020] [Medline: 27603721]
20. Bourne WM, Gebhardt BM, Sugar A, Meyer RF, Kaufman HE. The effect of splenectomy on corneal graft rejection. Invest Ophthalmol 1976 Jul;15(7):541-542. [Medline: 776894]
21. Hos D, Schaub F, Cursiefen C. Does anterior chamber-associated immune deviation (ACAID) play a role in posterior lamellar keratoplasty? case report of a splenectomized patient. BMC Ophthalmol 2019 May 02;19(1):100 [FREE Full text] [doi: 10.1186/s12886-019-1114-1] [Medline: 31046736]


## Abbreviations

**ACAID:** anterior chamber–associated immune deviation
**DMEK:** Descemet membrane endothelial keratoplasty

*Edited by T de Azevedo Cardoso; submitted 20.06.23; peer-reviewed by A Saxena, Y Wang; comments to author 19.01.24; revised version received 08.02.24; accepted 25.07.24; published 10.09.24.*

*Please cite as:*
*Kammrath Betancor P, Böhringer D, Maier P, Lapp T, Reinhard T*
*Splenectomy as a Risk Factor for Graft Rejection Following Endothelial Transplantation: Retrospective Study*
*Interact J Med Res 2024;13:e50106*
*URL: https://www.i-jmr.org/2024/1/e50106*
*doi:10.2196/50106*
*PMID:*

Original Paper

# Cardiovascular Comorbidities in COVID-19: Comprehensive Analysis of Key Topics

Rene Markovič[1,2], PhD; Luka Ternar[3], MD; Tim Trstenjak[3], MD; Marko Marhl[1,3,4], PhD; Vladimir Grubelnik[2], PhD

[1]Faculty of Natural Sciences and Mathematics, University of Maribor, Maribor, Slovenia

[2]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[3]Faculty of Medicine, University of Maribor, Maribor, Slovenia

[4]Faculty 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.

***(Interact J Med Res 2024;13:e55699)*** doi:10.2196/55699

**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 ventricles 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

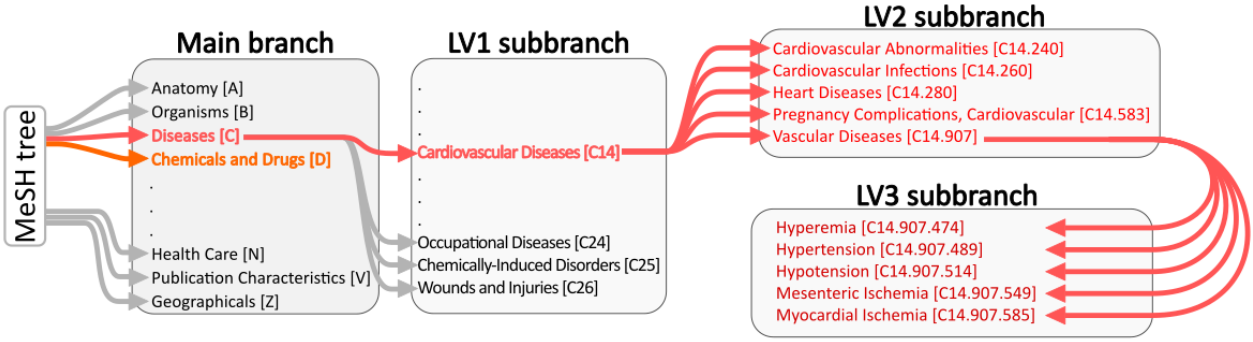
We conducted a search on PubMed [45] with specific search queries on CVD and SARS-CoV-2 and limited our search to articles published from the year 2000 onwards. Between January 1, 2000, and September 30, 2021, we collected all relevant entries in PubMed. From these entries, we selected only peer-reviewed 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.

**Figure 1.** A schematic representation of the hierarchical structure within the MeSH (Medical Subject Headings) tree, illustrating the organization from main branches to more specific sub-branches. The main branch example shown here includes categories like Anatomy [A], Diseases [C], and Chemicals and Drugs [D]. It details the progression from a main branch (Diseases [C]) to a LV1 sub-branch (Cardiovascular Diseases [C14]), to more refined LV2 and LV3 sub-branches, which specify narrower topics such as Heart Diseases [C14.280] and further down to Hypertension [C14.907.489] within the LV3 sub-branch. Each descriptor or topic is paired with a unique alphanumeric code that facilitates indexing and retrieval in medical databases. LV1: level 1; LV2: level 2; LV3: level 3.

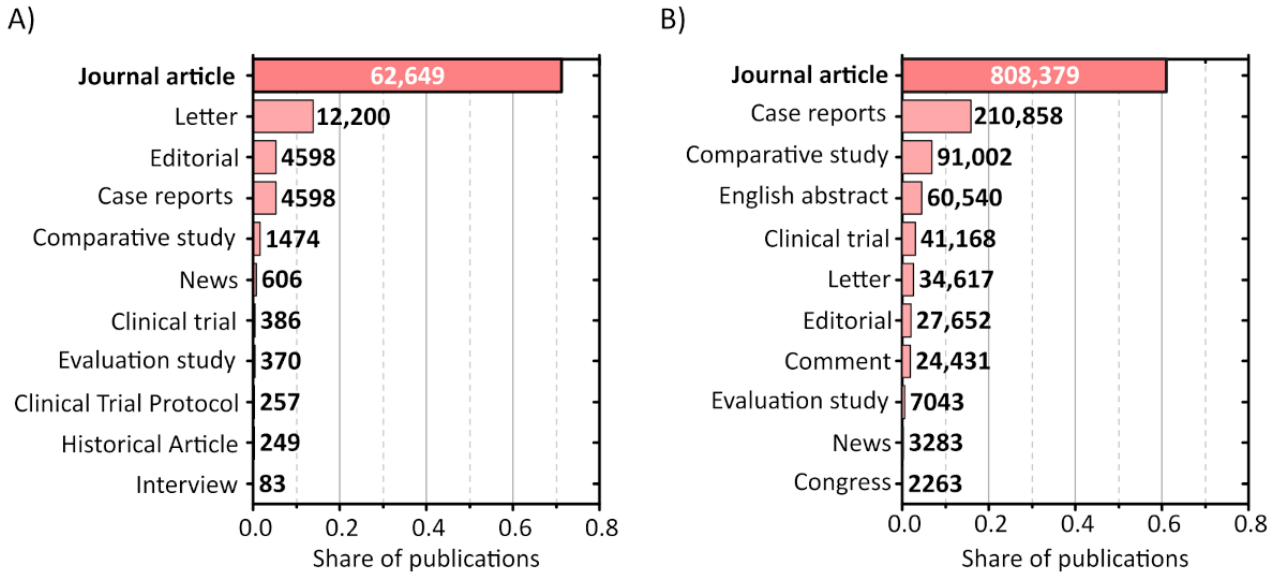



For our analysis, we developed a Python script capable of mining relevant publications from PubMed through their API. It extracts the MeSH descriptors associated with an article, along with its unique PMID, and translates a given MeSH descriptor code to the corresponding MeSH descriptor name. Since the 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.

**Figure 2.** Publication type prevalence in the (A) COVID-19 and (B) cardiovascular disease (CVD) data sets.




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

For each database, we performed statistical analyses using Python [46] and its associated libraries: pandas [47] for data manipulation, SciPy [48] for statistical calculations, and Matplotlib [49] for data visualization.

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:

$$\Delta f(MeSH) = f_{COVID}(MeSH) - f_{CVD}(MeSH) \quad \textbf{(1)}$$

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:

**(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

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, $f_{\mathrm{COVID}}(MeSH)$. 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.

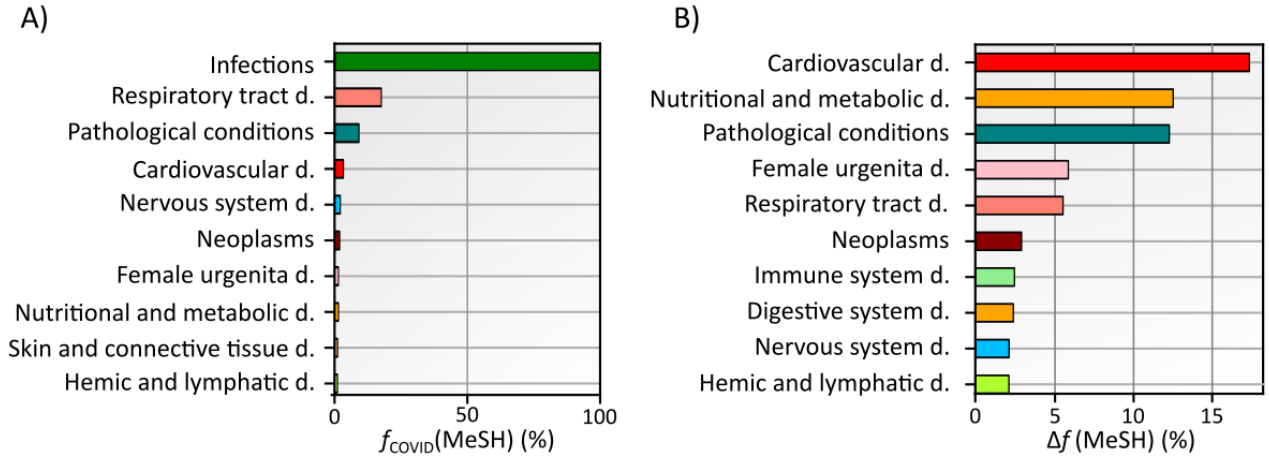



**Figure 4.** Most important level 2 (LV2) disease descriptor. Results are computed for (A) the entire COVID-19 data set and (B) for the subset of COVID-19 articles related to comorbidities. Each bar's color corresponds to a specific disease descriptor, as defined in Figure 3. MeSH: Medical Subject Headings.

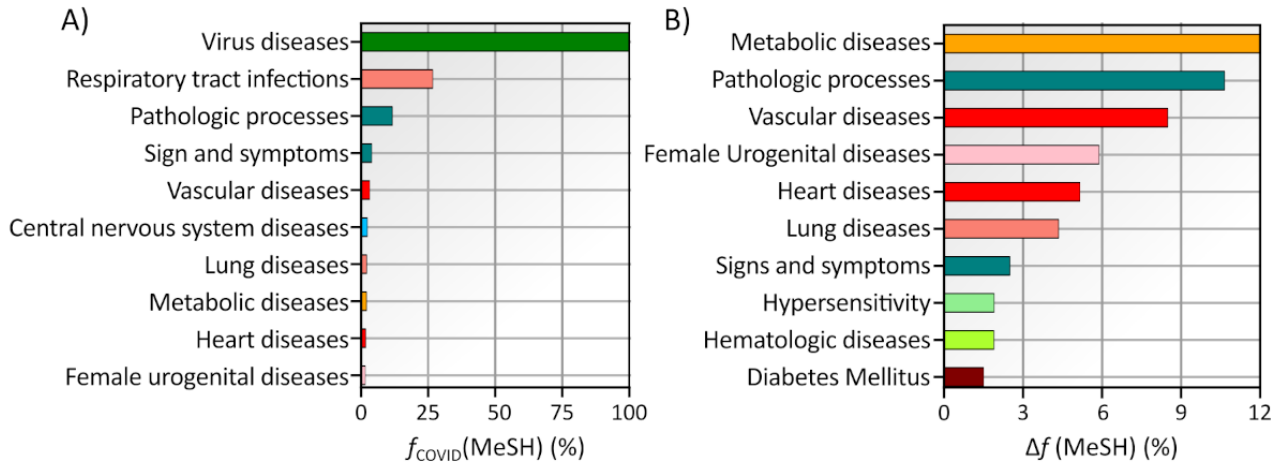


**Figure 5.** Ten most important level 3 (LV3) disease descriptors for COVID-19–related articles. Results are computed for (A) all LV3 disease descriptors, (B) only for LV3 disease descriptors originating from the "Nutritional and Metabolic Diseases," (C) only for the LV3 "Cardiovascular Diseases" branch, 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.

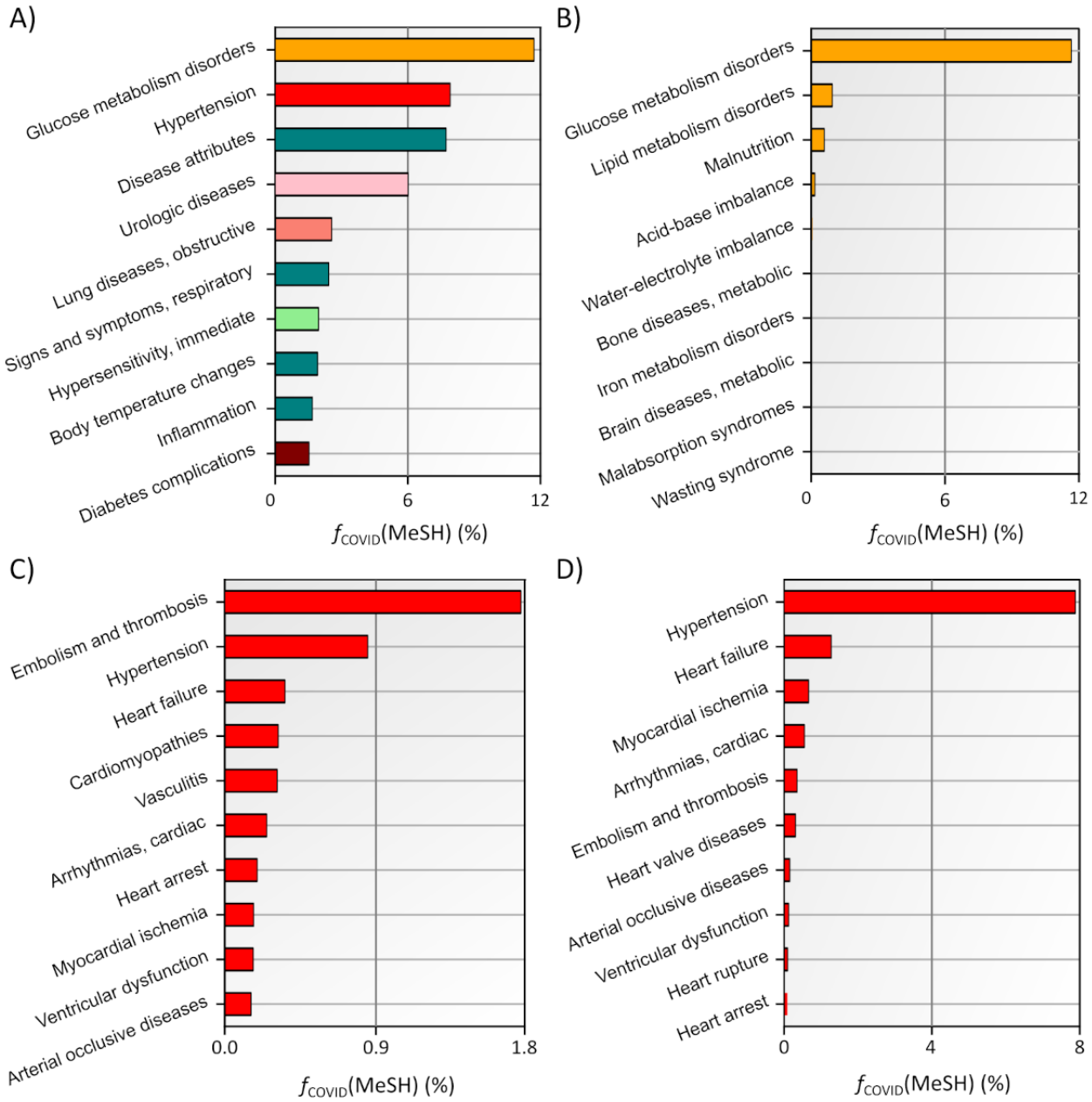
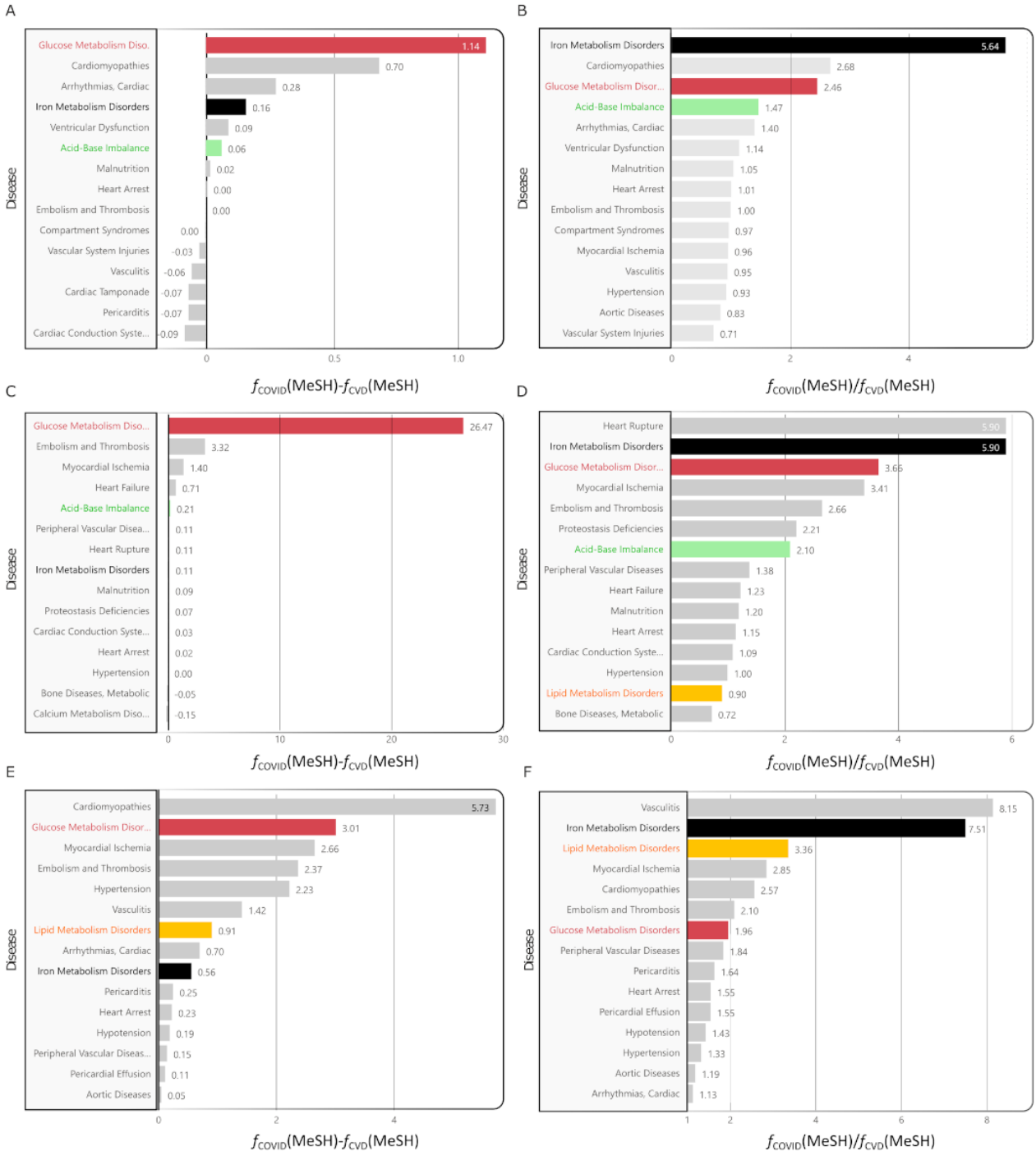


**Figure 6.** A comprehensive comparison of the absolute and relative changes in the frequency of level 3 (LV3) MeSH (Medical Subject Headings) descriptors in the COVID-19 database relative to the cardiovascular disease (CVD) database, focusing on 3 specific cardiovascular diseases: Embolism and Thrombosis (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 ($\Delta f$[MeSH]=0.11%) but a higher ratio ($R$[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 (with a substantial increase of $\Delta f$[MeSH]=26.47% and a ratio of $R$[MeSH]=3.66), and "Acid-Base Imbalance" is third ($\Delta f$[MeSH]=0.21% and $R$[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)[a].

| LV3[b] MeSH terms | Embolism and thrombosis, *P* value | Hypertension, *P* value | Heart failure, *P* value |
|---|---|---|---|
| Glucose Metabolism Disorders | <.001 | <.001 | <.05 |
| Iron Metabolism Disorders | <.01 | .525 | .116 |
| Acid-Base Imbalance | .103 | .42 | .585 |
| Lipid Metabolism Disorders | .94 | .53 | .05 |

[a]The *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.

[b]LV3: 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 CVD 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

1. Bradshaw RA, Stahl PD. On pandemics, pandemonium, and possibilities. FASEB Bioadv 2020;2(6):329-330 [FREE Full text] [doi: 10.1096/fba.2020-00023] [Medline: 32617518]
2. Herschhorn A, Haase AT. Science at its best in the time of the Covid-19 pandemic. ACS Infect Dis 2021;7(8):2209-2210. [doi: 10.1021/acsinfecdis.1c00269] [Medline: 34080826]
3. Forman R, Shah S, Jeurissen P, Jit M, Mossialos E. COVID-19 vaccine challenges: what have we learned so far and what remains to be done? Health Policy 2021 May;125(5):553-567 [FREE Full text] [doi: 10.1016/j.healthpol.2021.03.013] [Medline: 33820678]
4. Kim E, Erdos G, Huang S, Kenniston TW, Balmert SC, Carey CD, et al. Microneedle array delivered recombinant coronavirus vaccines: immunogenicity and rapid translational development. EBioMedicine 2020 May;55:102743 [FREE Full text] [doi: 10.1016/j.ebiom.2020.102743] [Medline: 32249203]
5. Leng L, Cao R, Ma J, Mou D, Zhu Y, Li W, et al. Pathological features of COVID-19-associated lung injury: a preliminary proteomics report based on clinical samples. Signal Transduct Target Ther 2020 Oct 15;5(1):240 [FREE Full text] [doi: 10.1038/s41392-020-00355-9] [Medline: 33060566]
6. Garima S, Singh N. Fatality in Covid 19: an overview of causes of death and organ involvement. Int J Adv Med 2020 Jun 22;7(7):1190-1193 [FREE Full text] [doi: 10.18203/2349-3933.ijam20202598]

7. Marhl M, Grubelnik V, Magdič M, Markovič R. Diabetes and metabolic syndrome as risk factors for COVID-19. Diabetes Metab Syndr 2020;14(4):671-677 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.013] [Medline: 32438331]
8. Nuzzi V, Del Mestre E, Degrassi A, Bromage DI, Manca P, Piper S, et al. Cardiovascular damage in COVID-19: what we know two years later. Curr Cardiol Rep 2022 Sep;24(9):1085-1091 [FREE Full text] [doi: 10.1007/s11886-022-01730-4] [Medline: 35751835]
9. Altay O, Mohammadi E, Lam S, Turkez H, Boren J, Nielsen J, et al. Current status of Covid-19 therapies and drug repositioning applications. iScience 2020;23(7):101303 [FREE Full text] [doi: 10.1016/j.isci.2020.101303] [Medline: 32622261]
10. Mangalmurti N, Hunter CA. Cytokine storms: understanding COVID-19. Immunity 2020 Jul 14;53(1):19-25 [FREE Full text] [doi: 10.1016/j.immuni.2020.06.017] [Medline: 32610079]
11. Mannucci PM. von Willebrand factor: a marker of endothelial damage? Arterioscler Thromb Vasc Biol 1998 Sep;18(9):1359-1362. [doi: 10.1161/01.atv.18.9.1359] [Medline: 9743222]
12. Vosko I, Zirlik A, Bugger H. Impact of COVID-19 on Cardiovascular Disease. Viruses 2023 Feb 11;15(2):508 [FREE Full text] [doi: 10.3390/v15020508] [Medline: 36851722]
13. Junaid K, Dawood N, Daood M, Randhawa FA, Yousaf MK, Ahmad MS. Prevalence, clinical characteristics, and clinical outcomes of new-onset diabetes mellitus among COVID-19 Patients in Developing and Developed Countries: a systematic review. J Coll Physicians Surg Pak 2023 Jun;33(6):691-699. [doi: 10.29271/jcpsp.2023.06.691] [Medline: 37300267]
14. Bielecka E, Sielatycki P, Pietraszko P, Zapora-Kurel A, Zbroch E. Elevated arterial blood pressure as a delayed complication following COVID-19—a narrative review. Int J Mol Sci 2024 Feb 02;25(3):1837 [FREE Full text] [doi: 10.3390/ijms25031837] [Medline: 38339115]
15. Shu H, Wen Z, Li N, Zhang Z, Ceesay BM, Peng Y, et al. COVID-19 and cardiovascular diseases: from cellular mechanisms to clinical manifestations. Aging Dis 2023 Dec 01;14(6):2071-2088 [FREE Full text] [doi: 10.14336/AD.2023.0314] [Medline: 37199573]
16. Datta PK, Liu F, Fischer T, Rappaport J, Qin X. SARS-CoV-2 pandemic and research gaps: understanding SARS-CoV-2 interaction with the ACE2 receptor and implications for therapy. Theranostics 2020;10(16):7448-7464 [FREE Full text] [doi: 10.7150/thno.48076] [Medline: 32642005]
17. Bourgonje AR, Abdulle AE, Timens W, Hillebrands J, Navis GJ, Gordijn SJ, et al. Angiotensin-converting enzyme 2 (ACE2), SARS-CoV-2 and the pathophysiology of coronavirus disease 2019 (COVID-19). J Pathol 2020;251(3):228-248 [FREE Full text] [doi: 10.1002/path.5471] [Medline: 32418199]
18. Ludwig S, Zarbock A. Coronaviruses and SARS-CoV-2: A Brief Overview. Anesth Analg 2020 Jul;131(1):93-96 [FREE Full text] [doi: 10.1213/ANE.0000000000004845] [Medline: 32243297]
19. Hamming I, Timens W, Bulthuis MLC, Lely AT, Navis GJ, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol 2004 Jun;203(2):631-637 [FREE Full text] [doi: 10.1002/path.1570] [Medline: 15141377]
20. Fernández-Ruiz I. ACE2 level as a marker of CVD. Nat Rev Cardiol 2020 Dec;17(12):759 [FREE Full text] [doi: 10.1038/s41569-020-00468-2] [Medline: 33082580]
21. Srichawla BS, Kipkorir V, Sekhon M. Clinical indications and outcomes of Impella devices for severe cardiogenic shock in COVID-19 patients: a systematic review. Ann Med Surg (Lond) 2023 Dec;85(12):6115-6122 [FREE Full text] [doi: 10.1097/MS9.0000000000001425] [Medline: 38098597]
22. Salabei JK, Asnake ZT, Ismail ZH, Charles K, Stanger GT, Abdullahi AH, et al. Covid-19 and the cardiovascular system: an update. Am J Med Sci 2022;364(2):139-147 [FREE Full text] [doi: 10.1016/j.amjms.2022.01.022] [Medline: 35151635]
23. Wibawa K, Nastiti KS, Meiviani SA, Syah PA, Suhendiwijaya S, Ariffudin Y. Review article: cardiac arrhythmia among hospitalized Covid-19 patients. Am Heart J 2023;32(1):8-18. [doi: 10.31762/ahj2332.0103]
24. Wang C, Hao W. Cardiac arrhythmia and immune response in Covid-19 patients. Allergol Immunopathol (Madr) 2023;51(4):63-70. [doi: 10.15586/aei.v51i4.883] [Medline: 37422781]
25. Vyas P, Mishra A, Parwani K, Patel I, Dhokia D, Amin T, et al. Significance of myocardial injury on in-hospital clinical outcomes of in-hospital and Covid-19 patients. J Cardiovasc Thorac Res 2023;15(2):93-97 [FREE Full text] [doi: 10.34172/jcvtr.2023.31614] [Medline: 37654814]
26. Patone M, Mei XW, Handunnetthi L, Dixon S, Zaccardi F, Shankar-Hari M, et al. Risks of myocarditis, pericarditis, and cardiac arrhythmias associated with COVID-19 vaccination or SARS-CoV-2 infection. Nat Med 2022 Feb;28(2):410-422 [FREE Full text] [doi: 10.1038/s41591-021-01630-0] [Medline: 34907393]
27. Wan EYF, Mathur S, Zhang R, Yan VKC, Lai FTT, Chui CSL, et al. Association of Covid-19 with short- and long-term risk of cardiovascular disease and mortality: a prospective cohort in UK Biobank. Cardiovasc Res 2023;119(8):1718-1727. [doi: 10.1093/cvr/cvac195] [Medline: 36652991]
28. Augustine R, Nayeem A, Salam SA, Augustine P, Dan P, Maureira P, et al. Increased complications of Covid-19 in people with cardiovascular disease: role of the renin-angiotensin-aldosterone system (RAAS) dysregulation. Chem Biol Interact 2022;351:109738 [FREE Full text] [doi: 10.1016/j.cbi.2021.109738] [Medline: 34740598]
29. Laragh JH. Atrial natriuretic hormone, the renin-aldosterone axis, and blood pressure-electrolyte homeostasis. N Engl J Med 1985;313(21):1330-1340. [doi: 10.1056/NEJM198511213132106] [Medline: 2932646]

30. Tikellis C, Thomas MC. Angiotensin-converting enzyme 2 (ACE2) is a key modulator of the renin angiotensin system in health and disease. Int J Pept 2012;2012:256294 [FREE Full text] [doi: 10.1155/2012/256294] [Medline: 22536270]
31. Deshotels MR, Xia H, Sriramula S, Lazartigues E, Filipeanu CM. Angiotensin II mediates angiotensin converting enzyme type 2 internalization and degradation through an angiotensin II type I receptor-dependent mechanism. Hypertension 2014;64(6):1368-1375 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.114.03743] [Medline: 25225202]
32. Nicin L, Abplanalp WT, Mellentin H, Kattih B, Tombor L, John D, et al. Cell type-specific expression of the putative SARS-CoV-2 receptor ACE2 in human hearts. Eur Heart J 2020;41(19):1804-1806 [FREE Full text] [doi: 10.1093/eurheartj/ehaa311] [Medline: 32293672]
33. Tucker NR, Chaffin M, Bedi KC, Papangeli I, Akkad AD, Arduini A, Human Cell Atlas Lung Biological Network, Human Cell Atlas Lung Biological Network Consortium Members. Myocyte-specific upregulation of in cardiovascular disease: implications for SARS-CoV-2-mediated myocarditis. Circulation 2020;142(7):708-710 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.120.047911] [Medline: 32795091]
34. Soler MJ, Batlle M, Riera M, Campos B, Ortiz-Perez JT, Anguiano L, et al. ACE2 and ACE in acute and chronic rejection after human heart transplantation. Int J Cardiol 2019;275:59-64. [doi: 10.1016/j.ijcard.2018.10.002] [Medline: 30314840]
35. Pinto BGG, Oliveira AER, Singh Y, Jimenez L, Gonçalves ANA, Ogava RLT, et al. ACE2 expression is increased in the lungs of patients with comorbidities associated with severe Covid-19. J Infect Dis 2020;222(4):556-563 [FREE Full text] [doi: 10.1093/infdis/jiaa332] [Medline: 32526012]
36. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med 2020 Apr;8(4):e21 [FREE Full text] [doi: 10.1016/S2213-2600(20)30116-8] [Medline: 32171062]
37. Guo J, Wei X, Li Q, Li L, Yang Z, Shi Y, et al. Single-cell RNA analysis on ACE2 expression provides insights into SARS-CoV-2 potential entry into the bloodstream and heart injury. J Cell Physiol 2020 Dec;235(12):9884-9894 [FREE Full text] [doi: 10.1002/jcp.29802] [Medline: 32510598]
38. Muñoz-Durango N, Fuentes CA, Castillo AE, González-Gómez LM, Vecchiola A, Fardella CE, et al. Role of the renin-angiotensin-aldosterone system beyond blood pressure regulation: molecular and cellular mechanisms involved in end-organ damage during arterial hypertension. Int J Mol Sci 2016 Jun 23;17(7):797 [FREE Full text] [doi: 10.3390/ijms17070797] [Medline: 27347925]
39. Gajjala PR, Sanati M, Jankowski J. Cellular and molecular mechanisms of chronic kidney disease with diabetes mellitus and cardiovascular diseases as its comorbidities. Front Immunol 2015;6:340 [FREE Full text] [doi: 10.3389/fimmu.2015.00340] [Medline: 26217336]
40. Amusa LB, Twinomurinzi H, Phalane E, Phaswana-Mafuya RN. Big data and infectious disease epidemiology: bibliometric analysis and research agenda. Interact J Med Res 2023 Mar 31;12:e42292 [FREE Full text] [doi: 10.2196/42292] [Medline: 36913554]
41. Zhang Y, Liu X, Qiao X, Fan Y. Characteristics and emerging trends in research on rehabilitation robots from 2001 to 2020: bibliometric study. J Med Internet Res 2023;25:e42901 [FREE Full text] [doi: 10.2196/42901] [Medline: 37256670]
42. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res 2020 Jul 29;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]
43. Saad RK, Abu Khudair S, El Rabbat M, Omar M, Al Nsour M, Khader Y, et al. Published research on COVID-19 in the Eastern Mediterranean region: bibliometric analysis. Interact J Med Res 2022 Jul 19;11(2):e38935 [FREE Full text] [doi: 10.2196/38935] [Medline: 35852846]
44. Chen Y, Cao B, Zhou Q, Liu Y, He Q, Zhao M. Bibliometric evaluation of 2020-2022 publications on COVID-19-related cardiovascular disease. Front Cardiovasc Med 2022;9:1070336 [FREE Full text] [doi: 10.3389/fcvm.2022.1070336] [Medline: 36712251]
45. PubMed - NCBI. URL: https://www.ncbi.nlm.nih.gov/pubmed/ [accessed 2020-04-23]
46. Python. Python Software Foundation. URL: https://www.python.org/ [accessed 2017-09-18]
47. McKinney W. Data structures for statistical computing in python. 2010 Presented at: Proceedings of the 9th Python in Science Conference; 2010; Austin, Texas p. 56-61. [doi: 10.25080/majora-92bf1922-00a]
48. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
49. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007 May;9(3):90-95. [doi: 10.1109/mcse.2007.55]
50. Adu-Amankwaah J, Mprah R, Adekunle AO, Ndzie Noah ML, Adzika GK, Machuki JO, et al. The cardiovascular aspect of COVID-19. Ann Med 2021 Dec;53(1):227-236 [FREE Full text] [doi: 10.1080/07853890.2020.1861644] [Medline: 33345626]
51. Zheng YY, Ma YT, Zhang JY, Xie X. COVID-19 and the cardiovascular system. Nat Rev Cardiol 2020 May;17(5):259-260 [FREE Full text] [doi: 10.1038/s41569-020-0360-5] [Medline: 32139904]
52. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res Cardiol 2020 May;109(5):531-538 [FREE Full text] [doi: 10.1007/s00392-020-01626-9] [Medline: 32161990]

53. Wu Z, McGoogan JM. Characteristics of and important lessons from the Coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. J Am Med Assoc 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
54. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. J Am Med Assoc 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
55. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
56. Driggin E, Madhavan MV, Bikdeli B, Chuich T, Laracy J, Biondi-Zoccai G, et al. Cardiovascular considerations for patients, health care workers, and health systems during the COVID-19 pandemic. J Am Coll Cardiol 2020 May 12;75(18):2352-2371 [FREE Full text] [doi: 10.1016/j.jacc.2020.03.031] [Medline: 32201335]
57. Lane HC, Marston HD, Fauci AS. Conducting clinical trials in outbreak settings: points to consider. Clin Trials 2016 Feb;13(1):92-95 [FREE Full text] [doi: 10.1177/1740774515618198] [Medline: 26768564]
58. Girelli D, Marchi G, Busti F, Vianello A. Iron metabolism in infections: focus on COVID-19. Semin Hematol 2021 Jul;58(3):182-187 [FREE Full text] [doi: 10.1053/j.seminhematol.2021.07.001] [Medline: 34389110]
59. Yilmaz N, Eren E, Öz C, Kalayci Z, Saribek F. Covid-19 and iron metabolism: traditional review. Turkiye Klin J Med Sci 2021;41(2):176-188. [doi: 10.5336/medsci.2021-81574]
60. Lv Y, Chen L, Liang X, Liu X, Gao M, Wang Q, et al. Association between iron status and the risk of adverse outcomes in COVID-19. Clin Nutr 2021 May;40(5):3462-3469 [FREE Full text] [doi: 10.1016/j.clnu.2020.11.033] [Medline: 33380357]
61. Sung HP, Hsu CY, Lee YH, Lin PL, Liao CT, Chung FP, et al. Iron deficiency in Taiwanese patients with heart failure and reduced ejection fraction. J Chin Med Assoc 2023 Aug 01;86(8):725-731. [doi: 10.1097/JCMA.0000000000000949] [Medline: 37314316]
62. Shamsi A, Cannata A, Piper S, Bromage DI, McDonagh TA. Treatment of iron deficiency in heart failure. Curr Cardiol Rep 2023 Jul;25(7):649-661 [FREE Full text] [doi: 10.1007/s11886-023-01889-4] [Medline: 37329419]
63. AlAayedi K. An audit of iron deficiency in hospitalised heart failure patients: a commonly neglected comorbidity. Cureus 2023 Jul;15(7):e41515 [FREE Full text] [doi: 10.7759/cureus.41515] [Medline: 37551218]
64. Park JJ, Yoon M, Cho HW, Lee SE, Choi JO, Yoo BS, et al. Iron deficiency in Korean patients with heart failure. J Korean Med Sci 2023 Jun 12;38(23):e177 [FREE Full text] [doi: 10.3346/jkms.2023.38.e177] [Medline: 37309696]
65. Miñana G, Santas E, de la Espriella R, Núñez E, Lorenzo M, Núñez G, et al. Right ventricular function and iron deficiency in acute heart failure. Eur Heart J Acute Cardiovasc Care 2021 May 25;10(4):406-414. [doi: 10.1093/ehjacc/zuaa028] [Medline: 33620455]
66. Jankowska EA, Tkaczyszyn M, Ponikowski P. Myocardial iron content in non-ischaemic cardiomyopathy: how much is known? Eur J Heart Fail 2020 Nov;22(11):2047-2048 [FREE Full text] [doi: 10.1002/ejhf.2023] [Medline: 33034152]
67. Zhang H, Jamieson K, Grenier J, Nikhanj A, Tang J, Nguyen Q, et al. Myocardial iron depletion exacerbates end-stage heart failure by promoting adverse remodeling and worsening mitochondrial function. Eur Heart J 2020;41(Supplement_2). [doi: 10.1093/ehjci/ehaa946.1144]
68. Poor HD. Pulmonary thrombosis and thromboembolism in COVID-19. Chest 2021 Oct;160(4):1471-1480 [FREE Full text] [doi: 10.1016/j.chest.2021.06.016] [Medline: 34153340]
69. Lippi G, Mullier F, Favaloro EJ. D-dimer: old dogmas, new (COVID-19) tricks. Clin Chem Lab Med 2023 Apr 25;61(5):841-850 [FREE Full text] [doi: 10.1515/cclm-2022-0633] [Medline: 35849562]
70. Kollias A, Kyriakoulis KG, Lagou S, Kontopantelis E, Stergiou GS, Syrigos K. Venous thromboembolism in COVID-19: a systematic review and meta-analysis. Vasc Med 2021 Aug;26(4):415-425 [FREE Full text] [doi: 10.1177/1358863X21995566] [Medline: 33818197]
71. Ali MAM, Spinler SA. COVID-19 and thrombosis: from bench to bedside. Trends Cardiovasc Med 2021 Apr;31(3):143-160 [FREE Full text] [doi: 10.1016/j.tcm.2020.12.004] [Medline: 33338635]
72. Makhoul E, Aklinski JL, Miller J, Leonard C, Backer S, Kahar P, et al. A review of COVID-19 in relation to metabolic syndrome: obesity, hypertension, diabetes, and dyslipidemia. Cureus 2022 Jul;14(7):e27438 [FREE Full text] [doi: 10.7759/cureus.27438] [Medline: 36051728]
73. Liu F, Long X, Zhang B, Zhang W, Chen X, Zhang Z. ACE2 expression in pancreas may cause pancreatic damage after SARS-CoV-2 infection. Clin Gastroenterol Hepatol 2020 Aug;18(9):2128-2130.e2 [FREE Full text] [doi: 10.1016/j.cgh.2020.04.040] [Medline: 32334082]
74. Bastin A, Shiri H, Zanganeh S, Fooladi S, Momeni Moghaddam MA, Mehrabani M, et al. Iron chelator or iron supplement consumption in COVID-19? The role of iron with severity infection. Biol Trace Elem Res 2022 Nov;200(11):4571-4581 [FREE Full text] [doi: 10.1007/s12011-021-03048-8] [Medline: 34825316]
75. Suriawinata E, Mehta KJ. Iron and iron-related proteins in COVID-19. Clin Exp Med 2023 Aug;23(4):969-991 [FREE Full text] [doi: 10.1007/s10238-022-00851-y] [Medline: 35849261]

76. Zhang Y, Chen L, Xuan Y, Zhang L, Tian W, Zhu Y, et al. Iron overload in hypothalamic AgRP neurons contributes to obesity and related metabolic disorders. Cell Rep 2024 Mar 26;43(3):113900 [FREE Full text] [doi: 10.1016/j.celrep.2024.113900] [Medline: 38460132]
77. Kalff H, Cario H, Holzhauer S. Iron deficiency anemia and thrombosis risk in children-revisiting an old hypothesis. Front Pediatr 2022;10:926925 [FREE Full text] [doi: 10.3389/fped.2022.926925] [Medline: 35979407]

## 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

*Edited by T de Azevedo Cardoso; submitted 21.12.23; peer-reviewed by L Zhi, D Uppal; comments to author 09.03.24; revised version received 22.03.24; accepted 23.05.24; published 24.07.24.*

*Please cite as:*
*Markovič R, Ternar L, Trstenjak T, Marhl M, Grubelnik V*
*Cardiovascular Comorbidities in COVID-19: Comprehensive Analysis of Key Topics*
*Interact J Med Res 2024;13:e55699*
*URL: https://www.i-jmr.org/2024/1/e55699*
*doi:10.2196/55699*
*PMID:*

Original Paper

# Analyzing Comorbidity Patterns in Patients With Thyroid Disease Using Large-Scale Electronic Medical Records: Network-Based Retrospective Observational Study

Yanqun Huang[1], PhD; Siyuan Chen[2], MSc; Yongfeng Wang[1], MSc; Xiaohong Ou[1], MSc; Huanhuan Yan[1], MSc; Xin Gan[1], MSc; Zhixiao Wei[2], PhD

[1]Department of Medical Equipment, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China

[2]Department of Nuclear Medicine, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, China

**Corresponding Author:**
Zhixiao Wei, PhD
Department of Nuclear Medicine
The First Affiliated Hospital of Guangxi Medical University
Number 6, Shuangyong Road
Nanning, Guangxi, 530021
China
Phone: 86 13517880896
Email: weizhixiao196493@126.com

## Abstract

**Background:** Thyroid disease (TD) is a prominent endocrine disorder that raises global health concerns; however, its comorbidity patterns remain unclear.

**Objective:** This study aims to apply a network-based method to comprehensively analyze the comorbidity patterns of TD using large-scale real-world health data.

**Methods:** In this retrospective observational study, we extracted the comorbidities of adult patients with TD from both private and public data sets. All comorbidities were identified using ICD-10 (International Classification of Diseases, 10th Revision) codes at the 3-digit level, and those with a prevalence greater than 2% were analyzed. Patients were categorized into several subgroups based on sex, age, and disease type. A phenotypic comorbidity network (PCN) was constructed, where comorbidities served as nodes and their significant correlations were represented as edges, encompassing all patients with TD and various subgroups. The associations and differences in comorbidities within the PCN of each subgroup were analyzed and compared. The PageRank algorithm was used to identify key comorbidities.

**Results:** The final cohorts included 18,311 and 50,242 patients with TD in the private and public data sets, respectively. Patients with TD demonstrated complex comorbidity patterns, with coexistence relationships differing by sex, age, and type of TD. The number of comorbidities increased with age. The most prevalent TDs were nontoxic goiter, hypothyroidism, hyperthyroidism, and thyroid cancer, while hypertension, diabetes, and lipoprotein metabolism disorders had the highest prevalence and PageRank values among comorbidities. Males and patients with benign TD exhibited a greater number of comorbidities, increased disease diversity, and stronger comorbidity associations compared with females and patients with thyroid cancer.

**Conclusions:** Patients with TD exhibited complex comorbidity patterns, particularly with cardiocerebrovascular diseases and diabetes. The associations among comorbidities varied across different TD subgroups. This study aims to enhance the understanding of comorbidity patterns in patients with TD and improve the integrated management of these individuals.

***(Interact J Med Res 2024;13:e54891)*** doi:10.2196/54891

**KEYWORDS**

thyroid disease; comorbidity patterns; prevalence; network analysis; electronic medical records

## Introduction

Thyroid disease (TD) is a prominent endocrine disorder and has become an increasing public health concern worldwide. In the United States, its prevalence rose from 2.54% in 1999-2002 to 5.05% in 2015-2018 [1]. Thyroid cancer (TC) is considered the 9th most common malignancy and the most common endocrine cancer, accounting for over 586,000 cases and 43,600 deaths annually [2]. In China, thyroid nodules (TNs) were the most common form of TD, affecting about 36.9% of the population in 2017 [3]. The incidence of TC in China is nearly 2 times the global average, with 11.3 versus 6.6 cases per 100,000 people [4].

TD is intricately linked to multiple diseases and health conditions [5,6]. A comprehensive understanding of the interactions and overlapping symptoms among these diseases can enhance clinicians' diagnostic accuracy and personalize treatment plans [7]. However, previous studies have mostly focused on the correlation between a single disease and a specific type of TD [6,8], neglecting the simultaneous consideration of multiple diseases. Consequently, the underlying patterns of TD's multiple comorbidities are far from fully elucidated.

Advancements in network theory offer fresh insights into understanding the intricate relationships among comorbidities. Recently, the phenotypic comorbidity network (PCN) has gained popularity in exploring associations and disparities across multiple diseases. In the PCN, coexisting diseases are represented as nodes, with edges indicating their connections. These edges can be assigned weights to reflect the frequency of their coexistence. The PCN uncovers hidden disease patterns, facilitating enhanced comorbidity risk assessments and future disease predictions for individuals. Recently, researchers have used the PCN to reveal comorbidity patterns for specific diseases such as diabetes [9], colorectal cancer [10], and heart failure [11].

Electronic medical records (EMRs), a typical real-world data set, are increasingly mined for valuable insights enriched with clinical history information to enhance decision-making [12]. As a crucial component of EMR, discharge diagnoses reflect patients' co-occurring diseases and health status during hospitalization, offering insights into comorbidity associations. These disease diagnoses are typically identified by standardized International Classification of Diseases (ICD) codes, which have proven effective in elucidating comorbidity patterns among diseases in prior studies [7,10,13].

In this study, we aimed to leverage a network-based method to systematically investigate comorbidity patterns in a general population of patients with TD, using disease codes from large-scale real-world EMR data from both private and public data sets. We hope to uncover previously recognized or unrecognized relationships among multiple comorbidities of TD.

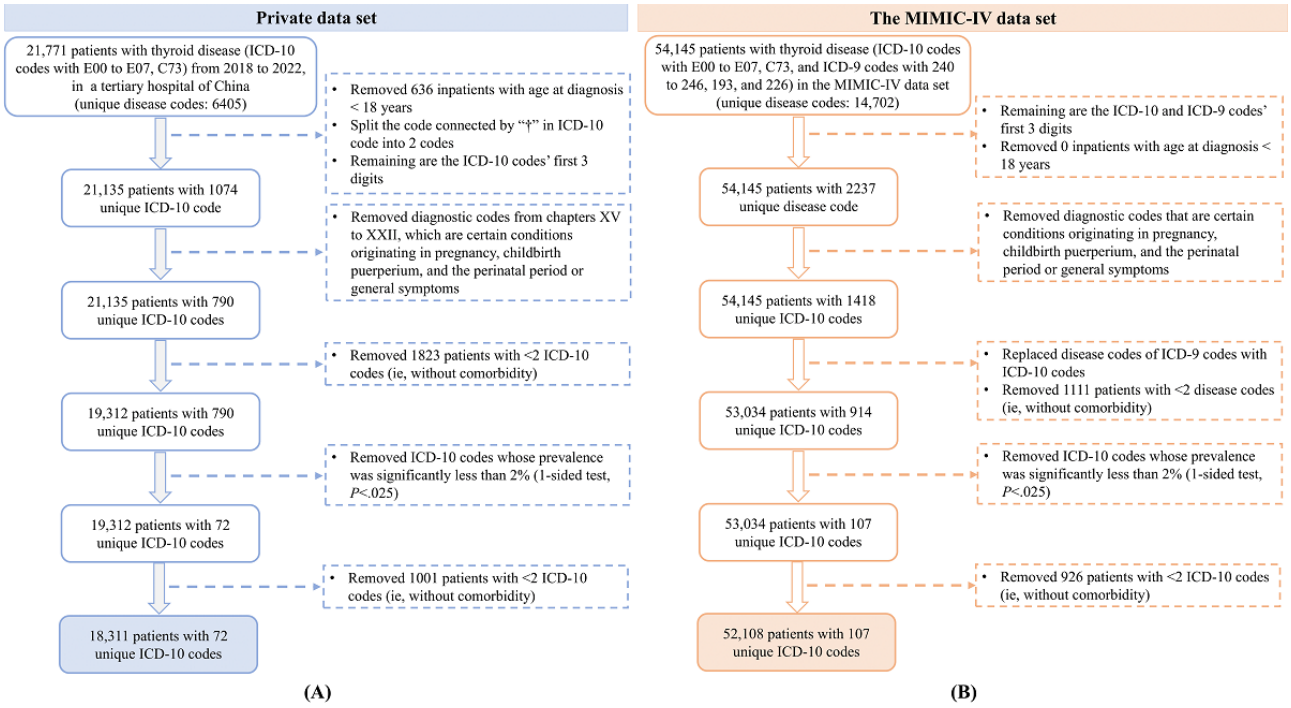
## Methods

### Study Populations

We conducted a retrospective cohort study using EMR data from both private and public data sets. Age, sex, and ICD codes were extracted. The private data set, covering 2018-2022, was obtained from a tertiary hospital in Guangxi, China, and included 21,771 hospitalizations and 6405 distinct ICD-10 (10th Revision) codes. Patients with TD were identified by ICD-10 codes E00-E07 (benign TD [BTD]) or C73 (TC). We treated each hospitalization record as an independent individual, considering that patients may have different health states and disease progression during different hospitalizations. The data cleaning process to identify the study cohort was as follows: (1) included adult patients aged ≥18 years; (2) excluded diagnostic codes from chapters XV to XXII, which pertain to pregnancy, childbirth, puerperium, perinatal conditions, and general symptoms; (3) excluded patient records with fewer than 2 ICD-10 codes; and (4) excluded rare diseases with a prevalence <2%. Ultimately, the study population comprised 18,311 patients in the private data set (Figure 1A).

Additionally, we validated our study using the Medical Information Mart for Intensive Care (MIMIC)-IV data set, a public clinical database containing over 73,000 admissions from 2008 to 2019. Diseases were initially coded using ICD-10 and ICD-9 (9th Revision) codes. To ensure consistency, we standardized all disease codes to 3-digit ICD-10 codes. Ultimately, we extracted data for 52,108 patients with TD from the MIMIC-IV data set (Figure 1B).

**Figure 1.** Flowchart of the study participants. ICD-10: International Classification of Diseases, 10th Revision.




## Ethical Considerations

The original data collection for this study was approved by the Medical Ethics Committee of the First Affiliated Hospital of Guangxi Medical University (approval number 2023-E592-01), and the Ethics Committee waived the requirement for a consent form. Informed consent from patients to use their data could not be obtained because the study is retrospective, observational, and population based. The analysis of the information was anonymized, with no variables that could identify individual patients. Additionally, we collected public clinical data from the MIMIC database after completing the web course (certification number 57439457). Given that the MIMIC-IV database consists of deidentified data and is publicly accessible, and as the collection of patient information and the creation of the research resource were reviewed by the Institutional Review Board at Beth Israel Deaconess Medical Center, which granted a waiver of informed consent and approved the data-sharing initiative [14], this study was not subject to specific ethical review.

## Comorbidity Prevalence

We stratified the population by sex (male or female) and disease type (TC or BTD). The prevalence of all diseases with ICD-10 codes and the 95% CI were calculated. We excluded diseases with a prevalence of <2% ($P$<.03, 1-sided test) in any subgroup to avoid including rare diseases. The lists of 72 and 107 diseases from the private and MIMIC-IV data sets are presented in Tables S1 and S2 in Multimedia Appendix 1, respectively. TDs, including TC (C73), nontoxic goiter (E04), hypothyroidism (E02 and E03), hyperthyroidism (E05), and other BTDs (E06), were excluded from the list of diseases, resulting in a final count of 67 and 102 comorbidities in the private and MIMIC-IV data sets, respectively. After comparing the comorbidity differences between the TC and BTD subgroups, we analyzed the comorbidity differences between patients with TC and those with specific typical BTDs (nontoxic goiter, E04; hypothyroidism, E02 and E03; and hyperthyroidism, E05). Subgroup details are presented in Figure S1 in Multimedia Appendix 1.

To evaluate the differences in comorbidity prevalence by sex and disease type, we calculated relative differences and used a $Z$-test for significance. The calculation for the relative difference by sex is based on equation (1) [10]:

$$d = (P_{male} - P_{female})/[(P_{male} + P_{female})/2] \quad \textbf{(1)}$$

where $P_{male}$ and $P_{female}$ represent the prevalences of comorbidities in males and females, respectively.

We considered comorbidity prevalence to be significantly different if the relative difference exceeded 0.1 [10] and the absolute difference was statistically significant. Among these, comorbidities with a ≥0.5-fold prevalence increase were deemed enriched. We used Spearman correlation to assess comorbidity prevalence trends with age and applied K-means clustering to group comorbidities by age-related prevalence.

## Phenotypic Comorbidity Network Construction

We constructed a PCN to capture the coexistence of multiple diseases. In the PCN, nodes represent disease codes (ICD-10 codes at 3 digits) that are connected by edges. Node sizes are proportional to disease prevalence, and node colors represent the ICD-10 categories. The cosine index was applied to quantify the comorbidity strength of coexisting diseases, taking into account the co-occurrence and prevalence of comorbidities, thereby minimizing the influence of sample size. We defined a cutoff value to detect comorbidity coexistence measured by the cosine index by assessing the relationship between the Pearson correlation and cosine index, where the number of significant coexisting comorbidities was equal in both networks as measured by the cosine index and Pearson correlation. The cosine index and Pearson correlation coefficient are defined in

equations (2) and (3), respectively, and the significance of was determined by performing a *t* test (2-tailed), calculated according to equation (4) [10].

where $N$ is the total number of patients with TD; and $n_a$, $n_b$, and $n_{ab}$ are the number of patients with disease a, disease b, and both diseases, respectively.

Four structural properties of the PCN were measured using network indices, including network density, degree, average degree of neighbors, and betweenness centrality [15]. Network density measures the compactness of a network by calculating the ratio of significant connections to all potential ones. A denser network indicates more connections among comorbidities. The degree of the PCN reflects its number of connections with other comorbidities. When a comorbidity is directly connected to others, those are referred to as neighbors. We calculated the average degree of neighbors to measure neighbor connectivity. Betweenness centrality represents the number of shortest paths between any 2 comorbidities. A high betweenness centrality indicates a greater likelihood of forming bridges between other comorbidities or serving as endpoints for many comorbidities [10].

To identify the most important comorbidity in the PCN, we applied the PageRank algorithm, which considers edge weights [16]. The PageRank algorithm calculates the importance of each node in a network by assigning ranks. Nodes that are connected to other nodes with a high rank receive a higher weight and are considered more central, while a higher PageRank value for a node indicates greater influence within the network. Comorbidities with the top 5 PageRank values were defined as the most important comorbidities in the PCN.

We constructed 4 separate PCNs for males, females, patients with TC, and patients with BTD. We then compared comorbidity strengths to measure disparities by sex and disease type. When a coexisting disease was unique or enriched (with a comorbidity strength of at least 0.05 higher than that of another) in a given subgroup, the difference was considered significant and was defined as an abundant connection [10].

Analyses and visualizations were performed using R 4.3.1 (R Foundation) and Python 3.7 (Python Foundation), leveraging the igraph and ggraph libraries for network visualization and property computation in R, and the pyecharts package for generating Sankey diagrams in Python.

## Results

### Characteristics of Patients With Thyroid Disease

In the private data set, the median age of patients with TD was 58 years, with 61.09% (11,187/18,311) being female and 6.25% (1144/18,311) having TC (Table 1). By contrast, in the MIMIC-IV data set, patients with TD were older, with a median age of 67 years; among them, 71.74% (37,384/52,108) were female and less than 0.93% (482/52,108) had TC. Patients with TD exhibited a high comorbidity burden, with a median number of comorbidities of 5 in the private data set and 7 in the MIMIC-IV data set. Males had more comorbidities than females, with the median number of comorbidities significantly higher in males than in females (6 vs 4, $P$<.001, and 8 vs 7, $P$<.001, in the private and MIMIC-IV data sets, respectively). Patients with TC were younger and had fewer comorbidities than patients with BTD. The median age of patients with TC was 16 and 11 years younger than that of patients with BTD, and the median number of comorbidities in patients with TC was 2 and 3 fewer than that in patients with BTD, in the private and MIMIC-IV data sets, respectively.

**Table 1.** Characteristics of patients with thyroid disease on the 2 data sets.

| Demographic and clinical factors | Private data set | | | | | MIMIC[a]-IV data set | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Overall | Male | Female | TC[b] | BTD[c] | Overall | Male | Female | TC | BTD |
| Number of patients | 18,311 | 7124 | 11,187 | 1144 | 17,167 | 52,108 | 14,724 | 37,384 | 482 | 51,626 |
| Age (years), median (IQR) | 58 (49-67) | 59 (51-68) | 57 (47-66) | 43 (34-53) | 59 (50-67) | 67 (55-78) | 68 (57-78) | 67 (55-79) | 56 (45-70) | 67 (55-79) |
| Number of comorbidities, median (IQR) | 5 (3-7) | 6 (4-8) | 4 (3-6) | 3 (2-4) | 5 (3-7) | 7 (5-10) | 8 (5-11) | 7 (5-10) | 4 (3-6) | 7 (5-10) |

[a]MIMIC: Medical Information Mart for Intensive Care.

[b]TC: thyroid cancer.

[c]BTD: benign thyroid disease.

The mean number of comorbidities increased with age in both males and females across both data sets (Figure 2A and 2C). In the private data set, the percentage of patients with TD decreased as the number of comorbidities increased, both overall and among females. By contrast, in the MIMIC-IV data set, the percentage of patients exhibited a trend of initially increasing and then decreasing with the number of comorbidities (Figure 2B and 2D).

The incidence of patients with different types of TDs varied between the 2 data sets (Figure 3). In the private data set, nontoxic goiter was the most common BTD, affecting 67.77% (12,410/18,311) of patients, whereas in the MIMIC-IV data set, hypothyroidism was prevalent in 86.36% (44,999/52,108) of

patients. In most patient subgroups, the percentage of patients decreased with the number of comorbidities in the private data set, while in the MIMIC-IV data set, the percentage of patients initially rose and then declined as the number of comorbidities increased (Figure 3B and 3E). However, across all TD subgroups within both data sets, the mean number of comorbidities increased with age, with Spearman correlations of 0.993 ($P$=.001) and 0.955 ($P$<.001) in the private and MIMIC-IV data sets, respectively (Figure 3C and 3F).

**Figure 2.** Comorbidity and age distributions in different subgroups of patients with thyroid disease. (A) Age-specific mean number of comorbidities and (B) frequency of patients per distinct number of comorbidities per patient on the private data set, respectively. Panels (C) and (D) are the counterparts of (A) and (B) on the MIMIC-IV data set, respectively. MIMIC: Medical Information Mart for Intensive Care.

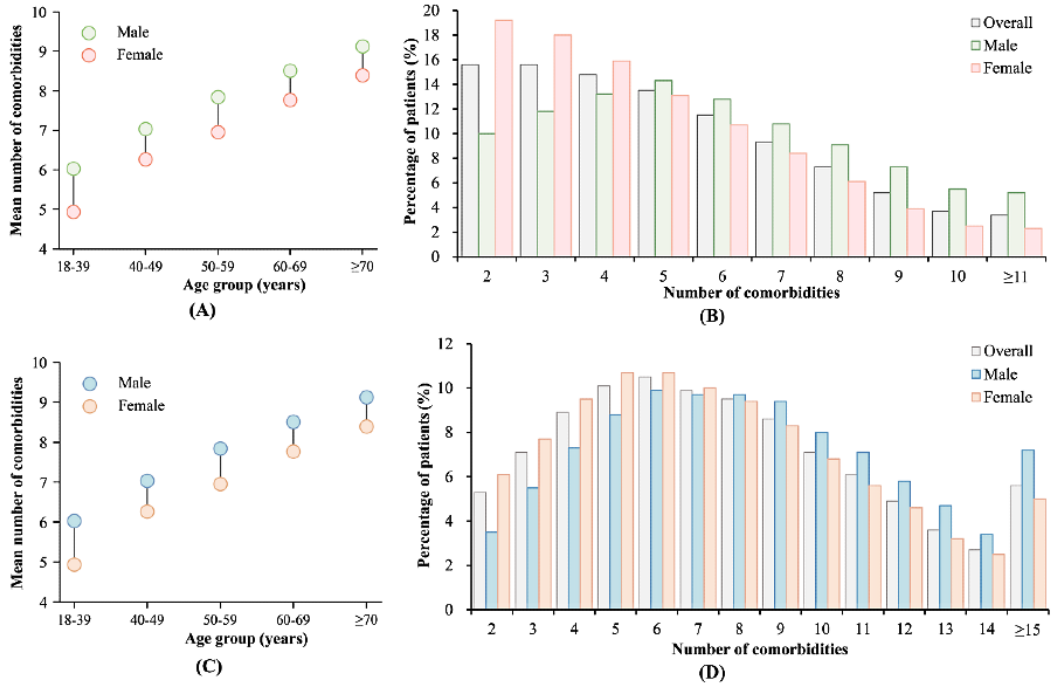
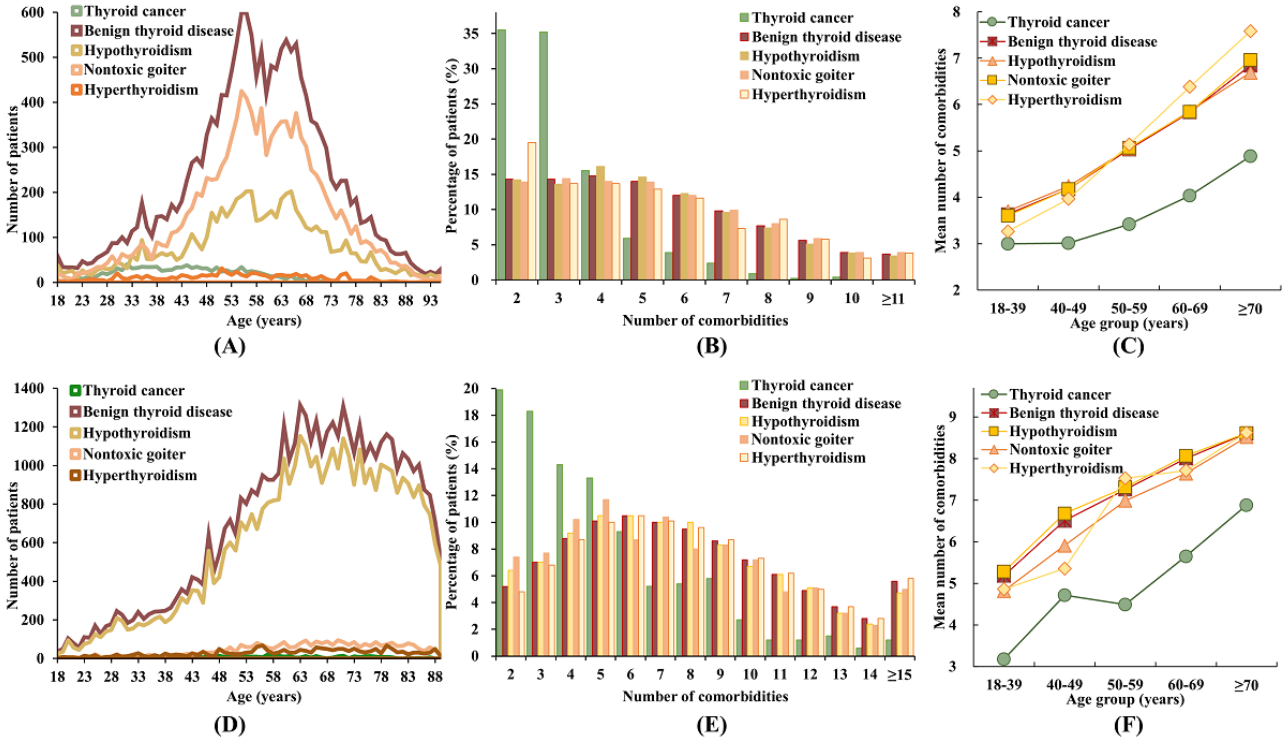



**Figure 3.** (A-C) The number of patients across different age groups, the percentage of patients with varying numbers of comorbidities, and the mean number of comorbidities for each thyroid disease subgroup, respectively, based on the private data set. (D-F) The corresponding data for patients from the MIMIC-IV data set. MIMIC: Medical Information Mart for Intensive Care.



### Comorbidity Prevalence and Differences Between Subgroups

Figure 4 displays the top 15 diseases with the highest prevalence. In the private data set, the most prevalent comorbidities were hypertension, liver disease, and diabetes, with prevalences of 35.14% (6435/18,311), 21.97% (4023/18,311), and 16.82% (3080/18,311), respectively. In the MIMIC-IV data set, hypertension and diabetes were also among the most common comorbidities, with prevalence rates of 42.13% (21,951/52,108) and 20.39% (10,627/52,108), respectively. Additionally, disorders of lipoprotein metabolism (E78) showed high prevalence, at 15.29% (2800/18,311) in the private data set and 44.37% (23,120/52,108) in the MIMIC-IV data set.

Among comorbidities with significant differences ($P$<.05) in prevalence, the differences were substantial, with some comorbidities enriched in specific subgroups (Table 2). In the private data set, diabetes (E11), cerebrovascular diseases (I63 and I69), heart diseases (I11, I20, I25, I48, and I50), chronic obstructive pulmonary disease (J43 and J44), renal diseases (N04, N18, N20, and N28), and gout (M10) were enriched in males. Conversely, systemic connective tissue disorders (M35), dorsopathies (M48 and M50), and disorders of bone density and structure (M80 and M81) were enriched in females.

Most comorbidities had higher prevalence and were even enriched in patients with BTD, including heart disease (I20, I25, and I50), cerebrovascular disease (I63, I65, and I67), hypertension (I10), diabetes (E11), liver disease (K74), gastritis and duodenitis (K29), and renal disease (N18, N20, and N08). Notably, malignant neoplasms were enriched in the TC group, including nasopharyngeal carcinoma (C11), bronchus and lung cancer (C34), and lymph node cancers (C77). Tables S3 and S4 in Multimedia Appendix 1 present the complete prevalence differences of enriched comorbidities in the subgroups of patients with TD from the private and MIMIC-IV data sets, respectively.

Age-specific comorbidities were clustered into 3 groups in the private data set and 4 groups in the MIMIC-IV data set (Figure 5). Most comorbidities' prevalence generally increased with age, although the rates varied by cluster. In the private data set, cluster 1 comprised 55 comorbidities with low, mostly stable prevalence. Cluster 2 included 11 comorbidities with moderate prevalence and high growth rates, such as atherosclerosis (I70), ischemic heart disease (I25), and diabetes (E11). Cluster 3 contained a single comorbidity (hypertension, I10) with the highest growth rate with age. In the MIMIC-IV data set, cluster 1 included 2 comorbidities (lipoprotein metabolism disorder, E78, and hypertension, I10) that showed significant age-related increases. Cluster 2 had 9 comorbidities with moderate increases, while clusters 3 and 4 comprised 22 and 70 comorbidities with low prevalence, respectively.

**Figure 4.** The top 15 comorbidities with the highest prevalence of patients with thyroid disease on the (A) private and (B) MIMIC-IV data sets. MIMIC: Medical Information Mart for Intensive Care.

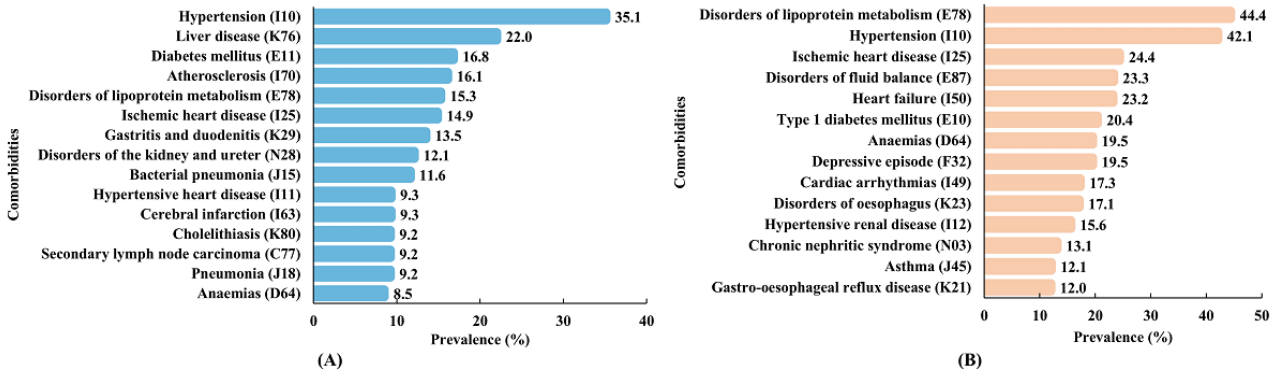


**Table 2.** Absolute prevalence differences of enrichment comorbidities in subgroups of patients with thyroid diseases on the private data set.

| ICD-10[a] code | Enrichment comorbidity | Crude prevalence (95% CI) | Absolute prevalence difference, value (95% CI) | |
|---|---|---|---|---|
| | | | Male – Female | TC[b] – BTD[c] |
| A49 | Bacterial infection | 2.7 (2.5 to 3.0) | N/A[d] | –2.6 (–3.0 to –2.2) |
| B66 | Fluke infections | 3.4 (3.1 to 3.6) | 5.1 (4.5 to 5.7) | –3.0 (–3.5 to –2.5) |
| C11 | Nasopharyngeal carcinoma | 1.7 (1.5 to 1.9) | 2.2 (1.8 to 2.6) | –1.7 (–2.0 to –1.4) |
| C22 | Malignant neoplasm of the liver and intrahepatic bile ducts | 1.7 (1.5 to 1.9) | 1.8 (1.4 to 2.2) | –1.8 (–2.0 to –1.6) |
| C34 | Bronchial and lung cancer | 7.8 (7.4 to 8.1) | 3.7 (2.9 to 4.5) | –7.5 (–8.1 to –6.9) |
| C50 | Malignant neoplasm of the breast | 2.4 (2.2 to 2.6) | –3.9 (–4.3 to –3.5) | –2.5 (–2.8 to –2.2) |
| C77 | Secondary lymph node carcinoma | 9.2 (8.8 to 9.7) | 4.0 (3.1 to 4.9) | 31.9 (29.0 to 34.8) |
| C78 | Secondary malignant neoplasm of the respiratory and digestive organs | 6.4 (6.1 to 6.8) | N/A | –3.0 (–4.1 to –1.9) |
| C79 | Secondary malignant neoplasm | 5.3 (5.0 to 5.6) | 2.9 (2.2 to 3.6) | –3.2 (–4.1 to –2.3) |
| C90 | Multiple myeloma and malignant plasma cell neoplasms | 1.7 (1.5 to 1.9) | 1.9 (1.5 to 2.3) | –1.8 (–2.0 to –1.6) |
| D18 | Hemangioma and lymphangioma | 3.1 (2.8 to 3.3) | N/A | –2.1 (–2.8 to –1.4) |
| D25 | Leiomyoma of the uterus | 3.3 (3.0 to 3.5) | –5.4 (–5.8 to –5.0) | –2.9 (–3.4 to –2.4) |
| D64 | Anemias | 8.5 (8.1 to 8.9) | N/A | –6.9 (–7.8 to –6.0) |
| E11 | Diabetes mellitus | 16.8 (16.3 to 17.4) | 7.2 (6.1 to 8.3) | –14.6 (–15.8 to –13.4) |
| E27 | Disorders of the adrenal gland | 3.3 (3.1 to 3.6) | 2.2 (1.6 to 2.8) | –3.4 (–3.8 to –3.0) |
| E55 | Vitamin D deficiency | 4.8 (4.5 to 5.2) | N/A | –5.0 (–5.4 to –4.6) |
| E77 | Disorders of glycoprotein metabolism | 4.4 (4.1 to 4.7) | N/A | –3.3 (–4.0 to –2.6) |
| E78 | Disorders of lipoprotein metabolism | 15.3 (14.8 to 15.8) | N/A | –12.5 (–13.7 to –11.3) |
| E79 | Disorders of purine and pyrimidine metabolism | 5.4 (5.1 to 5.7) | 3.8 (3.1 to 4.5) | –3.1 (–4.1 to –2.1) |
| E83 | Disorders of mineral metabolism | 1.6 (1.4 to 1.8) | N/A | 2.5 (1.3 to 3.7) |
| E87 | Disorders of fluid balance | 6.9 (6.6 to 7.3) | N/A | –3.5 (–4.7 to –2.3) |
| G31 | Degenerative diseases of the nervous system | 2.3 (2.0 to 2.5) | 1.1 (0.6 to 1.6) | –2.4 (–2.6 to –2.2) |
| G47 | Sleep disorders | 2.6 (2.4 to 2.9) | N/A | –2.7 (–3.0 to –2.4) |
| G63 | Polyneuropathy diseases | 3.3 (3.0 to 3.5) | N/A | –3.4 (–3.7 to –3.1) |
| I10 | Hypertension | 35.1 (34.5 to 35.8) | N/A | –25.2 (–27.2 to –23.2) |
| I11 | Hypertensive heart disease | 9.3 (8.9 to 9.7) | 4.2 (3.3 to 5.1) | –9.3 (–9.9 to –8.7) |
| I20 | Angina pectoris | 4.2 (3.9 to 4.5) | 3.5 (2.9 to 4.1) | –4.4 (–4.8 to –4.0) |
| I25 | Ischemic heart disease | 14.9 (14.4 to 15.5) | 8.1 (7.0 to 9.2) | –14.6 (–15.5 to –13.7) |
| I48 | Atrial fibrillation and flutter | 3.1 (2.9 to 3.4) | 1.5 (1.0 to 2.0) | –3.2 (–3.5 to –2.9) |
| I49 | Cardiac arrhythmias | 5.3 (5.0 to 5.6) | N/A | –5.3 (–5.8 to –4.8) |
| I50 | Heart failure | 7.4 (7.0 to 7.7) | 4.6 (3.8 to 5.4) | –7.8 (–8.2 to –7.4) |
| I63 | Cerebral infarction | 9.3 (8.9 to 9.7) | 4.2 (3.3 to 5.1) | –9.8 (–10.3 to –9.3) |
| I65 | Anterior cerebral artery stenosis and occlusion | 3.0 (2.7 to 3.2) | N/A | –3.2 (–3.5 to –2.9) |
| I67 | Cerebrovascular diseases | 4.2 (3.9 to 4.5) | N/A | –4.3 (–4.7 to –3.9) |
| I69 | Sequelae of cerebrovascular disease | 2.9 (2.7 to 3.1) | 2.3 (1.8 to 2.8) | –2.8 (–3.2 to –2.4) |
| I70 | Atherosclerosis | 16.1 (15.5 to 16.6) | N/A | –15.3 (–16.2 to –14.4) |
| I79 | Disorders of arteries, arterioles, and capillaries | 3.6 (3.3 to 3.9) | N/A | –3.9 (–4.2 to –3.6) |
| J15 | Bacterial pneumonia | 11.6 (11.1 to 12.0) | N/A | –11.8 (–12.4 to –11.2) |
| J18 | Pneumonia | 9.2 (8.8 to 9.7) | N/A | –5.8 (–7.0 to –4.6) |

| ICD-10[a] code | Enrichment comorbidity | Crude prevalence (95% CI) | Absolute prevalence difference, value (95% CI) | |
|---|---|---|---|---|
| | | | Male – Female | TC[b] – BTD[c] |
| J32 | Chronic sinusitis | 3.0 (2.8 to 3.3) | 1.6 (1.1 to 2.1) | –2.4 (–3.0 to –1.8) |
| J43 | Emphysema | 4.1 (3.8 to 4.3) | 5.1 (4.4 to 5.8) | –2.7 (–3.5 to –1.9) |
| J44 | Chronic obstructive pulmonary disease | 1.9 (1.7 to 2.1) | 3.2 (2.7 to 3.7) | –2.1 (–2.3 to –1.9) |
| J94 | Pleural conditions | 4.1 (3.8 to 4.4) | N/A | –3.9 (–4.4 to –3.4) |
| K29 | Gastritis and duodenitis | 13.5 (13.0 to 14.0) | N/A | –13.4 (–14.2 to –12.6) |
| K31 | Diseases of the stomach and duodenum | 2.9 (2.7 to 3.1) | N/A | –3.0 (–3.3 to –2.7) |
| K63 | Diseases of intestine | 2.6 (2.3 to 2.8) | N/A | –2.7 (–2.9 to –2.5) |
| K74 | Fibrosis and cirrhosis of the liver | 2.3 (2.1 to 2.5) | 1.7 (1.2 to 2.2) | –1.8 (–2.3 to –1.3) |
| K76 | Liver disease | 22.0 (21.4 to 22.6) | N/A | –17.4 (–18.9 to –15.9) |
| K80 | Cholelithiasis | 9.2 (8.7 to 9.6) | N/A | –7.7 (–8.6 to –6.8) |
| K82 | Diseases of the gallbladder | 3.3 (3.1 to 3.6) | N/A | –1.7 (–2.5 to –0.9) |
| M10 | Gout | 2.4 (2.2 to 2.6) | 4.6 (4.1 to 5.1) | –2.2 (–2.6 to –1.8) |
| M35 | Systemic involvement of connective tissue | 2.0 (1.8 to 2.2) | –2.5 (–2.8 to –2.2) | –2.0 (–2.3 to –1.7) |
| M48 | Spondylopathies | 4.8 (4.5 to 5.1) | –2.1 (–2.7 to –1.5) | –4.8 (–5.3 to –4.3) |
| M50 | Cervical disc disorders | 2.4 (2.2 to 2.6) | –1.3 (–1.7 to –0.9) | –2.4 (–2.7 to –2.1) |
| M51 | Intervertebral disc disorders | 4.0 (3.7 to 4.3) | N/A | –4.0 (–4.4 to –3.6) |
| M80 | Osteoporosis with pathological fracture | 1.9 (1.7 to 2.1) | –1.6 (–2.0 to –1.2) | –1.9 (–2.2 to –1.6) |
| M81 | Osteoporosis without pathological fracture | 7.1 (6.7 to 7.4) | –4.2 (–4.9 to –3.5) | –7.2 (–7.7 to –6.7) |
| N04 | Nephrotic syndrome | 3.0 (2.8 to 3.3) | 3.1 (2.5 to 3.7) | –3.2 (–3.5 to –2.9) |
| N08 | Glomerular disorders | 4.1 (3.8 to 4.3) | N/A | –4.2 (–4.6 to –3.8) |
| N18 | Chronic renal failure | 5.5 (5.1 to 5.8) | 4.5 (3.8 to 5.2) | –5.6 (–6.0 to –5.2) |
| N20 | Calculus of the kidney and ureter | 5.4 (5.0 to 5.7) | 3.0 (2.3 to 3.7) | –4.2 (–5.0 to –3.4) |
| N28 | Disorders of the kidney and ureter | 12.1 (11.6 to 12.6) | 6.0 (5.0 to 7.0) | –11.2 (–12.1 to –10.3) |
| N39 | Disorders of the urinary system | 3.0 (2.8 to 3.3) | –1.5 (–2.0 to –1.0) | –3.2 (–3.5 to –2.9) |
| N40 | Hyperplasia of the prostate | 7.5 (7.1 to 7.9) | 19.2 (18.3 to 20.1) | –7.7 (–8.2 to –7.2) |
| N60 | Benign mammary dysplasia | 1.5 (1.3 to 1.7) | –2.5 (–2.8 to –2.2) | N/A |

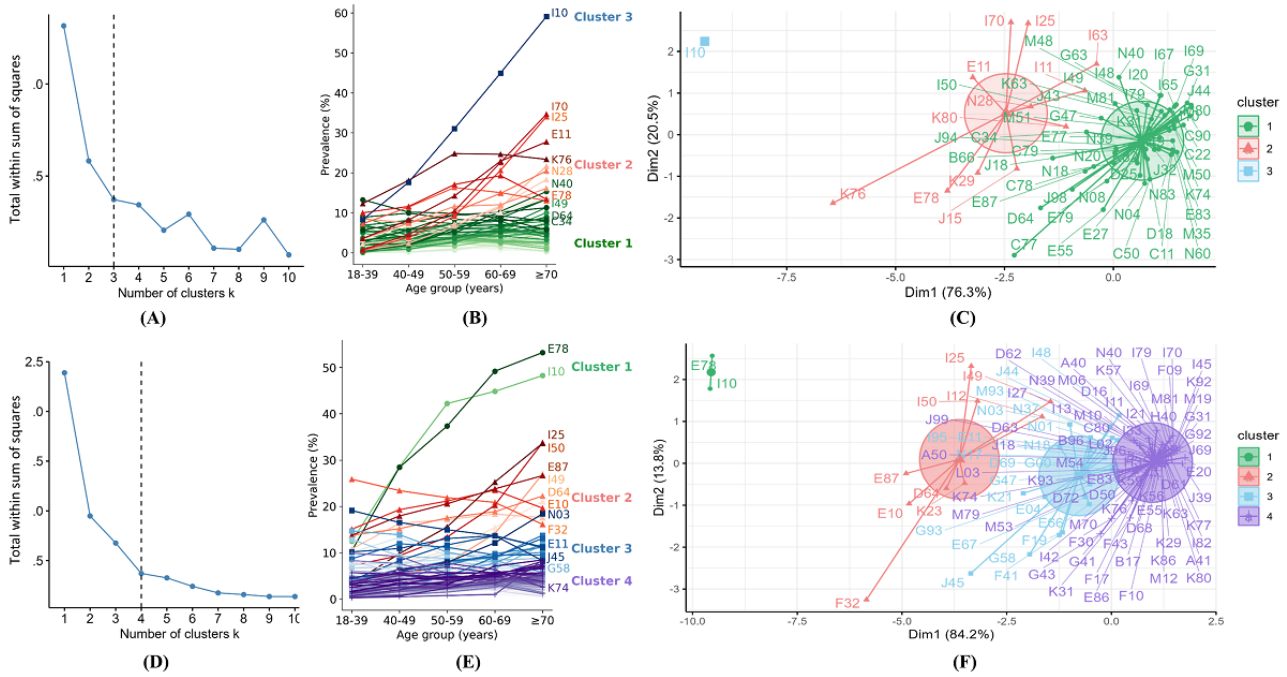
[a]ICD-10: International Classification of Diseases, 10th Revision.

[b]TC: thyroid cancer.

[c]BTD: benign thyroid disease.

[d]Not applicable, as the comorbidity was not enriched in this subgroup.

**Figure 5.** (A) and (D) The optimal number of clusters identified using the K-means clustering algorithm. (B) and (E) The age-specific prevalence of comorbidities within each cluster. (C) and (F) The cluster plots, where comorbidity prevalence in the 5 age groups (18-39, 40-49, 50-59, 60-69, and ≥70 years) has been reduced to 2 dimensions (x-axis and y-axis) through principal component analysis. Panels A-C are based on the private data set, while panels D-F correspond to the MIMIC-IV data set. A detailed list of ICD-10 codes is provided in Tables S1 and S2 in Multimedia Appendix 1. ICD-10: International Classification of Diseases, 10th Revision; MIMIC: Medical Information Mart for Intensive Care.



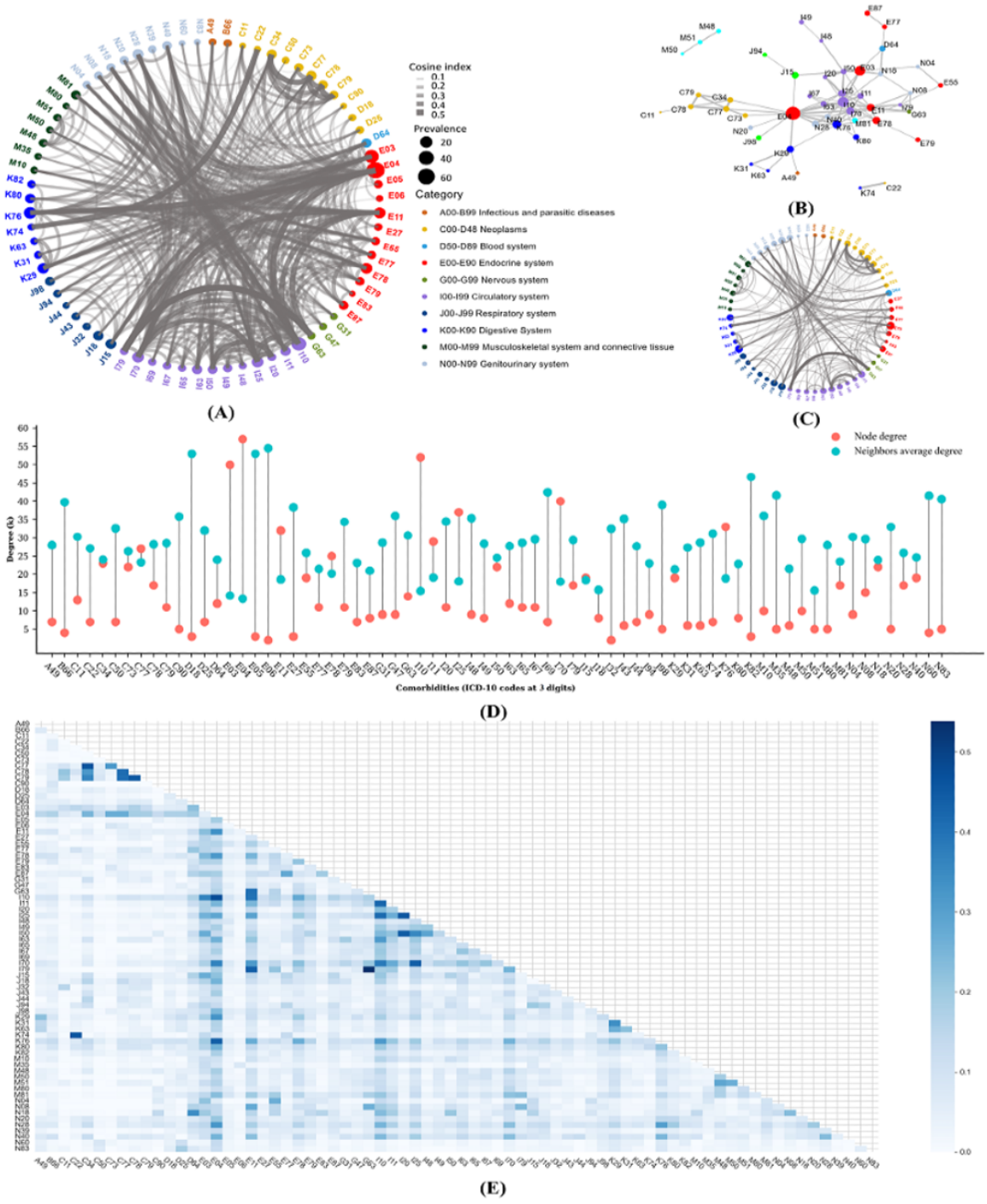

## Phenotypic Comorbidity Network in Patients With Thyroid Disease

In the private data set, the PCN of patients with TD contained 72 diseases and 492 disease pairs (Figure 6A). Each disease shared a median of 13.7 significant correlations with other comorbidities. Nontoxic goiter (E04), hypothyroidism (E03), and TC (C73) were prevalent TDs, with degrees of 50, 57, and 22, respectively (Figure 6D). Hypothyroidism (E03) was strongly linked to hypertension (I10), bacterial pneumonia (J15), heart failure (I50), ischemic heart disease (I25), and anemias (D64), with cosine indexes ranging from 0.23 to 0.36 (Figure 6A and 6E). By contrast, nontoxic goiter (E04) was associated with hypertension (I10; cosine index=0.47), liver disease (K76; cosine index=0.44), atherosclerosis (I70; cosine index=0.37), gastritis and duodenitis (K29; cosine index=0.33), and disorders of the renal and ureter (N28; cosine index=0.33), differing from those of hypothyroidism (E03). Nontoxic goiter (E04) and secondary lymph node carcinoma (C77) showed the highest correlation with TC (C73), with cosine indexes of 0.32 and 0.27, respectively, while other comorbidities had a cosine index of less than 0.06 with TC.

In the private data set, hypertension (I10), atherosclerosis (I70), ischemic heart disease (I25), diabetes (E11), and liver diseases (K76) were the top 5 comorbidities with high degrees (range 32-52) and betweenness centrality (range 47.1-385.2), reflecting their frequent coexistence and mediating role in disease correlations. Among the top 103 comorbid disease pairs with a cosine index of 0.20 or higher, some connections were within the same disease system, such as 50 and 12 connections within circulatory system diseases (ICD-10 codes starting with I) and malignant neoplasms (ICD-10 codes starting with C). When the diseases mentioned above (ie, E03, E04, I10, I70, I25, E11, and K76), which account for about 10% (7/72) of the diseases in the PCN, were removed, the scale of the PCN (Figure 6C) decreased significantly, with the number of significant comorbid disease pairs dropping by 57.5% (283/492).

In the MIMIC-IV data set, there were 107 diseases with 1819 disease links in the PCN, as shown in Figure S2 in Multimedia Appendix 1. Hypertension (I10), ischemic heart disease (I25), heart failure (I50), diabetes (E10 and E11), and esophageal disease (K23) were closely connected with TDs, with the highest cosine index values of 0.60, 0.48, 0.46, 0.42, and 0.39, respectively. Figures S3 and S4 in Multimedia Appendix 1 display the PCN for each subgroup in the private and MIMIC-IV data sets, respectively.

**Figure 6.** Phenotypic comorbidity network (PCN) in patients with thyroid disease (TD) on the private data set. (A) The PCN of patients with TD. Nodes represent comorbidities. The node size is proportional to the comorbidity prevalence in patients with TD and its color is used to identify the disease category. Link weights are proportional to the magnitudes of the cosine index which are >0. (B) The 103 edges in the PCN, for which the cosine index values are ≥0.20. (C) PCN of dropping the top 7 diseases with the highest degree and the highest betweenness centrality. (D) The degree distribution of the node and its neighbors. (E) Correlations measured by the cosine index between diseases. The disease codes are listed in Table S1 in Multimedia Appendix 1.




## Differences in Phenotypic Comorbidity Networks Between Subgroups

In both data sets, the comorbidity coexistence relationships were more complex in males than in females and in patients with BTD compared with those with TC (Table 3). In the private data set, males had 60 comorbidities and females had 63, with the median numbers of connections for comorbidities of 14.0 and 11.5, respectively. Patients with TC had fewer comorbidities (lower degrees) than those with BTD. As patients aged, the PCNs exhibited an increasing average degree of nodes and their neighboring nodes across the 5 age groups, indicating increasingly complex comorbidity coexistence relationships.

Figure 7 displays the top 10 important comorbidities with the highest PageRank values in different patient groups. In both data sets, cardiovascular and cerebrovascular diseases (colored in blue, ICD-10 codes with I), such as hypertension (I10 and I11) and ischemic heart disease (I25), along with endocrine and

metabolic diseases (colored in orange, ICD-10 codes with E), such as diabetes (E10 and E11) and lipoprotein metabolism disorder (E78), showed high PageRank values, indicating their significance in the PCN of patients with TD. Notably, secondary lymph node carcinoma (C77) exhibited a particularly high PageRank value within the PCN of patients with TD in the private data set, especially among patients with TC.

Figure 8 illustrates the abundant connections among patients with TD in the private data set. In the sex-specific PCN, there were 103 connections involving 55 comorbidities common to both sexes. Abundant connections in females were primarily linked to atrial fibrillation and flutter (I48), heart failure (I50), diseases of the stomach and duodenum (K29 and K31), and renal failure (N18). By contrast, abundant connections in males were mainly associated with various cardiocerebrovascular diseases (ICD-10 codes with I), malignant neoplasms (ICD-10 codes with C), bacterial pneumonia (J15), disorders of the ureter (N28), and cholelithiasis (K80). Differences in connections based on disease type were relatively minor, with 2 abundant connections in patients with TC and 5 in those without. The abundant connections in patients with TC were primarily linked to metabolic disorders (E78 and E79). By contrast, patients with BTD exhibited several abundant connections, including hypertension (I10), diabetes (E11), pneumonia (J18), and cholelithiasis (K80). The abundant connections within a specific subgroup of patients with TD from the MIMIC-IV data set are shown in Figure S5 in Multimedia Appendix 1.

**Table 3.** Phenotypic comorbidity network structures in subgroups of patients with thyroid disease.

| Subgroup | Nodes, n | | Density | | Degree, median (IQR) | | Degree of neighbors, median (IQR) | |
|---|---|---|---|---|---|---|---|---|
| | MIMIC[a]-IV | Private | MIMIC-IV | Private | MIMIC-IV | Private | MIMIC-IV | Private |
| All | 102 | 72 | 0.321 | 0.192 | 30.0 (20.0-46.0) | 9.0 (6.0-17.5) | 49.2 (44.0-54.9) | 28.3 (23.1-34.4) |
| **Sex** | | | | | | | | |
| Female | 90 | 60 | 0.312 | 0.240 | 27.0 (18.0-40.0) | 11.5 (6.0-16.0) | 43.0 (36.7-46.3) | 26.7 (21.3-30.7) |
| Male | 92 | 63 | 0.229 | 0.304 | 17.0 (10.8-27.8) | 14.0 (10.0-22.0) | 38.7 (33.2-43.0) | 32.9 (28.8-39.4) |
| **Disease subgroup** | | | | | | | | |
| TC[b] | 28 | 15 | 0.087 | 0.419 | 1.5 (1.0-2.0) | 4.0 (3.5-5.5) | 24.2 (16.5-31.0) | 11.0 (8.8-12.2) |
| BTD[c] | 100 | 69 | 0.338 | 0.198 | 30.0 (21.5-46.0) | 10.0 (6.0-16.0) | 48.4 (42.9-52.7) | 28.3 (23.6-34.6) |
| Nontoxic goiter | 81 | 65 | 0.111 | 0.217 | 6.0 (2.2-9.8) | 10.0 (6.0-15.0) | 31.2 (24.7-46.7) | 30.9 (22.2-34.0) |
| Hypothyroidism | 100 | 61 | 0.329 | 0.237 | 29.0 (20.0-43.0) | 11.0 (7.0-17.0) | 47.0 (41.4-51.8) | 28.2 (23.3-33.7) |
| Hyperthyroidism | 79 | 39 | 0.098 | 0.208 | 4.0 (2.0-9.2) | 5.0 (3.0-8.0) | 30.8 (21.6-46.1) | 23.0 (18.6-29.8) |
| **Age group (years)** | | | | | | | | |
| 18-39 | 57 | 46 | 0.115 | 0.133 | 6.0 (2.0-9.0) | 4.0 (3.0-6.0) | 11.9 (10.0-14.3) | 21.1 (15.8-29.7) |
| 40-49 | 71 | 55 | 0.127 | 0.143 | 7.0 (4.0-15.2) | 5.0 (3.0-7.0) | 17.5 (14.6-20.1) | 27.0 (22.6-33.0) |
| 50-59 | 85 | 66 | 0.171 | 0.207 | 12.0 (6.0-21.0) | 9.5 (6.0-14.0) | 25.7 (23.0-30.1) | 33.7 (26.7-39.1) |
| 60-69 | 92 | 65 | 0.206 | 0.265 | 17.0 (9.0-25.5) | 12.0 (8.0-19.0) | 34.6 (29.2-39.5) | 35.7 (28.9-42.6) |
| ≥70 | 89 | 59 | 0.265 | 0.353 | 22.0 (10.0-35.0) | 15.0 (11.0-22.5) | 39.2 (33.9-43.5) | 38.2 (32.0-41.2) |

[a]MIMIC: Medical Information Mart for Intensive Care.

[b]TC: thyroid cancer.

[c]BTD: benign thyroid disease.

**Figure 7.** Sankey diagrams of the top 10 important comorbidities of patients with thyroid disease in subgroups on the 2 data sets. The links between the subgroups and comorbidities were proportional to the PageRank values. The higher the PageRank value, the stronger the connection, and the more significant the comorbidity. MIMIC: Medical Information Mart for Intensive Care.

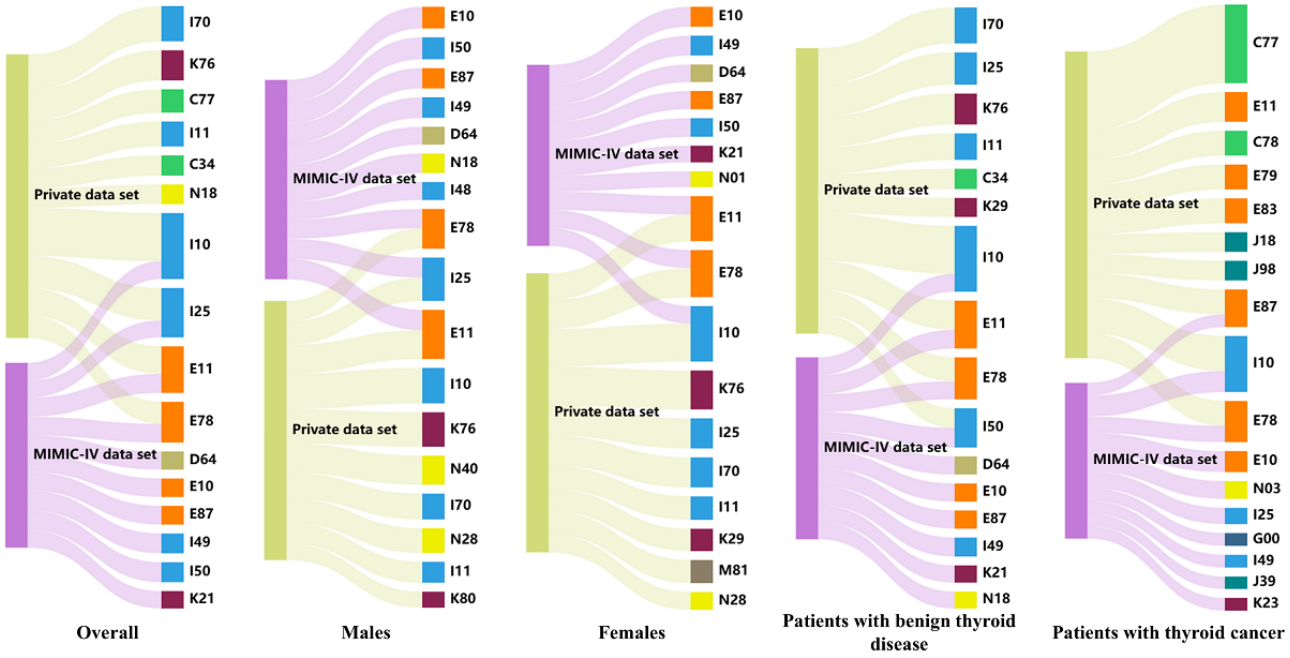
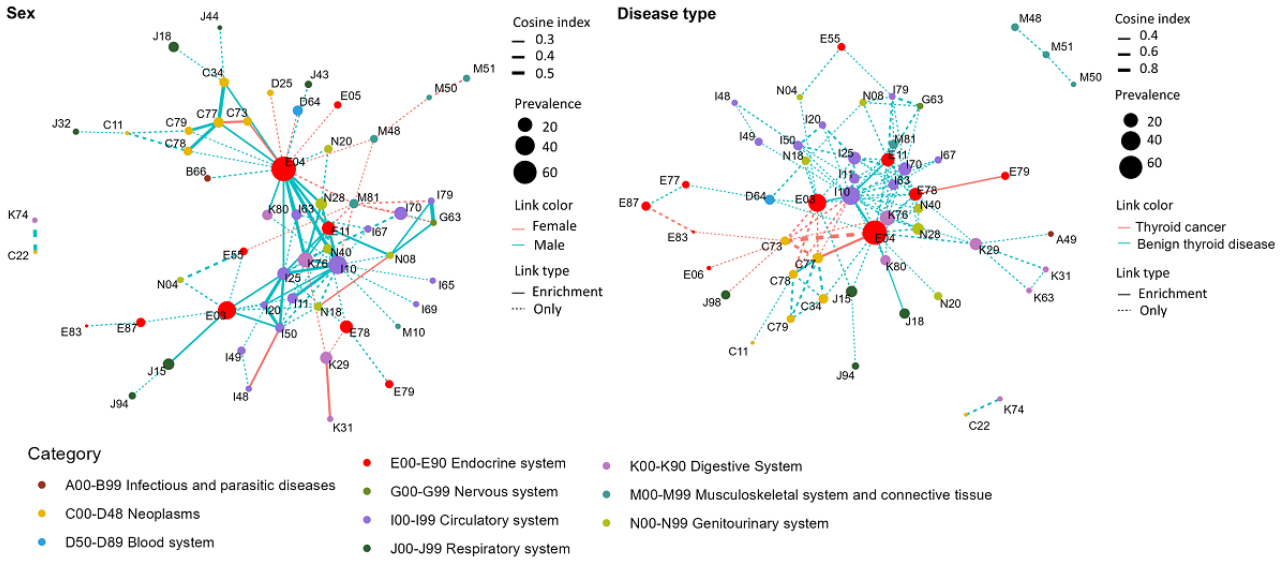



**Figure 8.** Abundant connections by sex and disease type in patients with thyroid disease on the private data set. Except for those common to both subgroups (link weight difference, ie, cosine index <0.05), disease pairs were identified as abundant edges in 1 subgroup, including enrichment (solid lines) or only occurrence (dotted lines).




## Discussion

### Principal Findings

In this population-based cohort study, we used a network-based approach to reveal the comprehensive comorbidity patterns in patients with TD, highlighting differences across sex, age, and TD types. Comorbidities were prevalent among patients with TD, with variations in prevalence and complex coexistence relationships influenced by sex, age, and the type of TD. Nontoxic goiter, hypothyroidism, hyperthyroidism, and TC were identified as the most common TDs. The top 5 comorbidities with the highest prevalence were hypertension, liver disease, diabetes, atherosclerosis, and disorders of lipoprotein metabolism. The number of comorbidities increased with age across all patients with TD and subgroups. Males and patients with BTD demonstrated higher rates of comorbidity, greater disease diversity, and stronger associations compared with females and patients with TC.

In this study, hypertension emerged as the most significant comorbidity, affecting 35.14% (6435/18,311) of patients with TD in the private data set. A previous study has demonstrated a positive correlation between blood pressure and thyroid-stimulating hormone levels [17]. Additionally, both blood pressure and dyslipidemia have been strongly associated with TNs [18]. Consistent with previous studies [5,19-22], we found a strong link between TD and diabetes. Specific thyroid

disorders, such as TNs [22], hypothyroidism [19], thyroid dysfunction [5], and thyroid disorders [20], have been linked to diabetes. Furthermore, thyroid function significantly impacts the cardiovascular system [23]. Both hyperthyroidism and hypothyroidism are associated with increased cardiovascular morbidity and mortality, and abnormal thyroid function elevates the risk of coronary heart disease [6,24]. Our findings also revealed high PageRank values for ischemic heart diseases within the PCN of patients with TD. Additionally, thyroid hormones regulate metabolism, linking metabolic syndrome and thyroid dysfunction to significant morbidity and mortality [23]. We also observed that metabolic disorders, such as disorders of lipoprotein metabolism, were closely associated with patients with TD. Moratalla et al [8], exploring the comorbidity network of hypothyroidism, showed strong links to thyroid/larynx cancer, hyperthyroidism, anemia, and goiter. We similarly found associations with nasopharyngeal carcinoma, secondary respiratory cancers, anemia, hyperthyroidism, and nontoxic goiter. However, we revealed that hypothyroidism was strongly associated with hypertension and pneumonia, which was unreported in the study by Moratalla et al [8].

In this study, females exhibited a higher TD burden than males, especially among younger individuals. Similar findings have been widely reported in several studies [22,25]. Consistent with previous studies [22,25], TDs were more prevalent in females than males [1], and the incidence of TC was significantly higher among females, with a ratio of 2.36 in 2017 [26]. This could be attributed to factors such as physiology, pregnancy, and estrogen exposure [27,28]. Estrogen, which is higher in females, can lead to thyroid gland changes, increasing the risk of TDs [27]. Therefore, greater attention should be paid to thyroid gland examination in females.

Further, our findings revealed sex-specific comorbidity patterns in patients with TD. Males, despite lower TD incidence, tended to have more comorbidities than females, with 31 enriched disease pairs in males' PCN compared with only 5 in females on the private data set. Moratalla et al [8] also found that comorbidity networks showed more and stronger connections in males than in females. Notably, cancers, especially malignant neoplasms of lymph nodes, bronchus and lung, and respiratory and digestive organs, demonstrated enriched connections with TC in the PCN of males. Moreover, despite evidence of higher primary TC incidence among females [26,29], males faced a greater risk of developing second primary TC compared with females [30-32]. For instance, males had a higher risk of head and neck cancer than females [30]. Our results confirmed that males were more likely to have concurrent cancers other than TC than females. Besides, males exhibited enrichment in multiple cardiocerebrovascular diseases, while females displayed fewer heart disease associations. Understanding these sex-specific comorbidity patterns aids in developing tailored treatment plans for patients with TD.

Our findings showed that older patients with TD, regardless of sex or TD type, tend to have more comorbidities, which increase with age. This aligns with previous research, including Stenholm et al [33], who reported a rise in diseases and physical difficulties with age, and Xu et al [22], who found a similar trend in TNs' prevalence. Therefore, enhanced nursing and care for older adult patients with TD is crucial to reduce comorbidities, improve quality of life, and ease medical burdens.

Notably, the average age of patients with TC in this study was 44 years, the same for both genders. A previous study [34] also found that 45-year-old women are the most prone to papillary thyroid carcinoma, while another study [26] reported peak ages of 50-69 years for males and 15-49 years for females. However, Huang et al [2] found that the incidence of TC increased in populations aged under 40 years in China, with males showing a higher growth rate (18.6 vs 13.3). Some researchers suggested that young patients may delay seeking treatment due to underestimating the cancer risk [35]. We speculated that high workload and mental stress in young people may cause endocrine issues and hormone imbalances, contributing to TC's younger trend. Given the link between neuropsychological functions and thyroid health [36,37], extra care is needed for young patients with TD to prevent cancer.

Besides, we observed that patients with TC, despite having a relatively low median comorbidity count (3 and 4 on the private and MIMIC-IV data sets, respectively), exhibited severe comorbidities such as cancer, heart failure, cerebral infarction, and pneumonia. This severity may stem from the aggressive nature of cancer and its tendency to metastasize, leading to life-threatening complications. Prior studies have shown higher risks of papillary TC following other neoplasms, particularly renal and breast cancers, as well as leukemias/lymphomas [38,39]. Similarly, Lian et al [40] found that subclinical hypothyroidism was common in patients with nasopharyngeal carcinoma within a year and that thyroid volume is a risk factor for radiation-induced hypothyroidism. Given these findings, heightened vigilance is necessary to prevent cancer progression and deterioration in patients with TC.

From a methodological perspective, we used cosine similarity and Pearson correlation indices to quantify node connectivity, considering disease prevalence and minimizing sample size biases. Cosine similarity effectively captured significant comorbidities in this study, as well as in previous analyses of comorbidities associated with colorectal cancer [10] and idiopathic cardiomyopathy [41]. Other similarity metrics, such as the Jaccard index [42], which quantifies the overlap between disease sets, and the observed-to-expected ratio, which considers the observed prevalence of disease pairs relative to the expected prevalence [7], were utilized in constructing cardiovascular comorbidity networks [7,42]. Additionally, various network properties were used to computationally identify key comorbidities. We applied the PageRank algorithm, which considers the number and weight of edges, the number of neighbor edges, and the centrality of neighbors, to assess the importance of comorbidities. In the comorbidity networks of both data sets in this study, the comorbidities selected by PageRank were similar, including hypertension and diabetes. PageRank has also been utilized in constructing comorbidity networks for patients with colorectal cancer [10], general patients [16], and hypothyroidism patients [8]. Additionally, other properties, such as degree [11] and degree centrality [42], have been used to identify important comorbidities and their connections.

## Limitations

This study has several limitations. First, apart from the MIMIC-IV data set, we collected data and conducted analyses solely in a general tertiary hospital in China, which may introduce selection bias, particularly concerning geographical and population factors. Utilizing multicenter data could yield more representative and robust results. Second, our analysis focused solely on the coexistence of comorbidities without considering their causal relationships, and many network properties remain unexplained in clinical or biological contexts. Some diseases may share common pathophysiological mechanisms or causes, leading to mutual influence. In some cases, one disease may promote the occurrence and progression of another. Therefore, further analysis is needed to explore the deeper and more detailed causal relationships between thyroid disorders and their comorbidities. Third, relying solely on ICD-10 codes for disease identification may not fully capture all cases related to thyroid disorders, potentially leading to missing information. Additionally, some diagnoses may be subject to overdiagnosis or underdiagnosis due to the complexity of clinical issues. Certain diseases may not be frequently coded, and broad categorization may obscure differences among heterogeneous conditions. Finally, we conducted only a cross-sectional analysis and did not consider the temporality and sequence of comorbidities when constructing comorbidity profiles. These patterns may change as diseases progress. By collecting more data over longer periods from multiple hospitals, a more detailed analysis could provide deeper insights into the comorbidity patterns of patients with thyroid disorders.

## Conclusions

This data-driven exploration of comorbidity patterns among all patients with thyroid disorder and their subgroups aims to provide valuable clinical and foundational insights into the comorbidities associated with thyroid disorders. It enhances our understanding of thyroid disorders as a whole and may inform the development of more effective and integrated therapeutic strategies for patients. Additionally, the network-based methodology utilized in this study has the potential to uncover comorbidity patterns in other diseases. In the future, we plan to incorporate a wider variety of data to construct more comprehensive networks. By extracting features from these networks, we aim to leverage medical knowledge to develop machine learning–based prediction models for patients with thyroid disorders.

## Acknowledgments

This study was supported by the Outstanding Young Doctoral Program Research Initiation Fund of the First Affiliated Hospital of Guangxi Medical University (grant 202302). We are grateful to Professor Hui Chen (The Capital Medical University, Beijing, China) for her methodological suggestions on our study.

## Data Availability

The MIMIC-IV data sets used in this study are available online [43]. However, the private data set generated or analyzed in this study is not publicly available as these are restricted by the Medical Ethics Committee of the First Affiliated Hospital of Guangxi Medical University.

## Authors' Contributions

YH was responsible for conceptualization, data curation, methodology, and writing the original draft. SC contributed to resources and validation. YW handled data curation and resources. XO and HY contributed to data curation. XG provided resources. ZW was involved in conceptualization, resources, writing, review, and editing.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional data analysis.
[DOCX File , 5490 KB - ijmr_v13i1e54891_app1.docx ]

## References

1. Zhang X, Wang X, Hu H, Qu H, Xu Y, Li Q. Prevalence and trends of thyroid disease among adults, 1999-2018. Endocr Pract 2023 Nov;29(11):875-880. [doi: 10.1016/j.eprac.2023.08.006] [Medline: 37619827]
2. Huang J, Ngai C, Deng Y, Pun C, Lok V, Zhang L, NCD Research Group, Global Health, Association of Pacific Rim Universities (APRU). Incidence and mortality of thyroid cancer in 50 countries: a joinpoint regression analysis of global trends. Endocrine 2023 May;80(2):355-365. [doi: 10.1007/s12020-022-03274-7] [Medline: 36607509]
3. Li Y, Jin C, Li J, Tong M, Wang M, Huang J, et al. Prevalence of Thyroid Nodules in China: A Health Examination Cohort-Based Study. Front Endocrinol 2021 May 26;12:676144 [FREE Full text] [doi: 10.3389/fendo.2021.676144]

4. Xie L, Wang S, Qian Y, Jia S, Wang J, Li L, et al. Increasing gap between thyroid cancer incidence and mortality in urban Shanghai, China: an analysis spanning 43 years. Endocr Pract 2021 Nov;27(11):1100-1107 [FREE Full text] [doi: 10.1016/j.eprac.2021.06.002] [Medline: 34119680]
5. Biondi B, Kahaly G, Robertson R. Thyroid dysfunction and diabetes mellitus: two closely associated disorders. Endocr Rev 2019 Jun 01;40(3):789-824 [FREE Full text] [doi: 10.1210/er.2018-00163] [Medline: 30649221]
6. Moon S, Kim M, Yu J, Yoo H, Park Y. Subclinical hypothyroidism and the risk of cardiovascular disease and all-cause mortality: a meta-analysis of prospective cohort studies. Thyroid 2018 Sep;28(9):1101-1110 [FREE Full text] [doi: 10.1089/thy.2017.0414]
7. Zhou D, Wang L, Ding S, Shen M, Qiu H. Phenotypic disease network analysis to identify comorbidity patterns in hospitalized patients with ischemic heart disease using large-scale administrative data. Healthcare (Basel) 2022 Jan 01;10(1):80 [FREE Full text] [doi: 10.3390/healthcare10010080] [Medline: 35052244]
8. Moratalla-Navarro F, Moreno V, López-Simarro F, Aguado A. MorbiNet study: hypothyroidism comorbidity networks in the adult general population. The Journal of Clinical Endocrinology and Metabolism 2021 Mar 8;106(3):e1179-e1190 [FREE Full text] [doi: 10.1210/clinem/dgaa927]
9. Hossain M, Uddin S. Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes. Expert Systems with Applications 2021 Feb;164:113918 [FREE Full text] [doi: 10.1016/j.eswa.2020.113918]
10. Qiu H, Wang L, Zhou L, Wang X. Comorbidity patterns in patients newly diagnosed with colorectal cancer: network-based study. JMIR Public Health Surveill 2023 Sep 05;9:e41999 [FREE Full text] [doi: 10.2196/41999] [Medline: 37669093]
11. Lanzer J, Valdeolivas A, Pepin M, Hund H, Backs J, Frey N, et al. A network medicine approach to study comorbidities in heart failure with preserved ejection fraction. BMC Med 2023 Jul 24;21(1):267 [FREE Full text] [doi: 10.1186/s12916-023-02922-7] [Medline: 37488529]
12. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29 [FREE Full text] [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
13. Mu X, Wang W, Wu F, Jiang Y, Ma L, Feng J. Comorbidity in older patients hospitalized with cancer in Northeast China based on hospital discharge data. Int J Environ Res Public Health 2020 Oct 31;17(21):8028 [FREE Full text] [doi: 10.3390/ijerph17218028] [Medline: 33142785]
14. Johnson A, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023 Jan 03;10(1):1 [FREE Full text] [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
15. Freeman L. Centrality in social networks conceptual clarification. Social Networks 1978 Jan;1(3):215-239 [FREE Full text] [doi: 10.1016/0378-8733(78)90021-7]
16. Amell A, Roso-Llorach A, Palomero L, Cuadras D, Galván-Femenía I, Serra-Musach J, et al. Disease networks identify specific conditions and pleiotropy influencing multimorbidity in the general population. Sci Rep 2018 Oct 29;8(1):15970 [FREE Full text] [doi: 10.1038/s41598-018-34361-3] [Medline: 30374096]
17. Wang J, Wang C, Pei D, Lai C, Chen Y, Wu C, et al. Association between thyroid function and metabolic syndrome in elderly subjects. J Am Geriatr Soc 2010 Aug 1;58(8):1613-1614 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.02998.x] [Medline: 20942887]
18. Liang Y, Li X, Wang F, Yan Z, Sang Y, Yuan Y, et al. Detection of thyroid nodule prevalence and associated risk factors in southwest China: a study of 45,023 individuals undergoing physical examinations. DMSO 2023 Jun 8;Volume 16:1697-1707 [FREE Full text] [doi: 10.2147/dmso.s412567]
19. Chen H, Wu T, Jap T, Lu R, Wang ML, Chen RL, et al. Subclinical hypothyroidism is a risk factor for nephropathy and cardiovascular diseases in type 2 diabetic patients. Diabet Med 2007 Dec;24(12):1336-1344. [doi: 10.1111/j.1464-5491.2007.02270.x] [Medline: 17941864]
20. Ghimire S, Sangroula P, Deo R, Ghimire S, Dhonju K. Spectrum of thyroid disorders in patients with type-2 diabetes mellitus. J Nepal Health Res Counc 2023 Jul 20;20(4):922-927 [FREE Full text] [doi: 10.33314/jnhrc.v20i4.4314] [Medline: 37489678]
21. Yasuda T, Kaneto H, Kuroda A, Yamamoto T, Takahara M, Naka T, et al. Subclinical hypothyroidism is independently associated with albuminuria in people with type 2 diabetes. Diabetes Res Clin Pract 2011 Dec;94(3):e75-e77. [doi: 10.1016/j.diabres.2011.08.019] [Medline: 21925759]
22. Xu L, Zeng F, Wang Y, Bai Y, Shan X, Kong L. Prevalence and associated metabolic factors for thyroid nodules: a cross-sectional study in Southwest of China with more than 120 thousand populations. BMC Endocr Disord 2021 Aug 28;21(1):175 [FREE Full text] [doi: 10.1186/s12902-021-00842-2] [Medline: 34452638]
23. Meng Z, Liu M, Zhang Q, Liu L, Song K, Tan J, et al. Gender and age impacts on the association between thyroid function and metabolic syndrome in Chinese. Medicine (Baltimore) 2015 Dec;94(50):e2193 [FREE Full text] [doi: 10.1097/MD.0000000000002193] [Medline: 26683929]
24. Cooper D, Biondi B. Subclinical thyroid disease. The Lancet 2012 Mar 24;379(9821):1142-1154 [FREE Full text] [doi: 10.1016/s0140-6736(11)60276-6]

25. Yang H, Zhong Y, Lv W, Zhang F, Yu H. Association of adiposity with thyroid nodules: a cross-sectional study of a healthy population in Beijing, China. BMC Endocr Disord 2019 Oct 16;19(1):102 [FREE Full text] [doi: 10.1186/s12902-019-0430-z] [Medline: 31619235]
26. Deng Y, Li H, Wang M, Li N, Tian T, Wu Y, et al. Global burden of thyroid cancer from 1990 to 2017. JAMA Netw Open 2020 Jun 01;3(6):e208759 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8759] [Medline: 32589231]
27. Li J, Liu Y, Jin J, Shi Q, Ji Y, Zhang B, et al. Study on clinicopathological features and risk factors of patients with multiple primary breast cancers and thyroid disease. Mediators Inflamm 2023 Apr 6;2023:3133554 [FREE Full text] [doi: 10.1155/2023/3133554] [Medline: 37152372]
28. Derwahl M, Nicula D. Estrogen and its role in thyroid cancer. Endocrine Related Cancer 2014 Jul 22;21(5):T273-T283 [FREE Full text] [doi: 10.1530/erc-14-0053]
29. Aupérin A. Epidemiology of head and neck cancers: an update. Curr Opin Oncol 2020 May 1;32(3):178-186 [FREE Full text] [doi: 10.1097/CCO.0000000000000629] [Medline: 32209823]
30. Hussein M, Mueller L, Issa P, Haidari M, Trinh L, Toraih E, et al. Sexual disparity and the risk of second primary thyroid cancer: a paradox. Gland Surg 2023 Apr 28;12(4):432-441 [FREE Full text] [doi: 10.21037/gs-22-411] [Medline: 37200932]
31. Sung H, Hyun N, Leach C, Yabroff K, Jemal A. Association of first primary cancer with risk of subsequent primary cancer among survivors of adult-onset cancers in the United States. JAMA 2020 Dec 22;324(24):2521-2535 [FREE Full text] [doi: 10.1001/jama.2020.23130] [Medline: 33351041]
32. Kim C, Bi X, Pan D, Chen Y, Carling T, Ma S, et al. The risk of second cancers after diagnosis of primary thyroid cancer is elevated in thyroid microcarcinomas. Thyroid 2013 May 1;23(5):575-582 [FREE Full text] [doi: 10.1089/thy.2011.0406] [Medline: 23237308]
33. Stenholm S, Westerlund H, Head J, Hyde M, Kawachi I, Pentti J, et al. Comorbidity and functional trajectories from midlife to old age: the Health and Retirement Study. J Gerontol A Biol Sci Med Sci 2015 Mar 1;70(3):332-338 [FREE Full text] [doi: 10.1093/gerona/glu113] [Medline: 25060316]
34. Lam A. Papillary thyroid carcinoma: current position in epidemiology, genomics, and classification. Methods Mol Biol 2022 Jan 1;2534:1-15 [FREE Full text] [doi: 10.1007/978-1-0716-2505-7_1] [Medline: 35670964]
35. Iqbal A, Azhar S, Ibrahim N, Kharaba Z, Iqbal M, Khan S, et al. Thyroid cancer risk factors and Pakistani University students' awareness towards its preventive practice. J Oncol Pharm Pract 2021 Apr 1;27(3):570-578 [FREE Full text] [doi: 10.1177/1078155220925166] [Medline: 32437225]
36. Yang R, Zhu F, Yue Y, Lu X, Zhu P, Li Z, et al. Association between thyroid function and psychotic symptoms in adolescents with major depressive disorder: a large sample sized cross-sectional study in China. Heliyon 2023 Jun 27;9(6):e16770 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e16770] [Medline: 37303557]
37. Tang R, Wang J, Yang L, Ding X, Zhong Y, Pan J, et al. Subclinical hypothyroidism and depression: a systematic review and meta-analysis. Front Endocrinol (Lausanne) 2019 Jun 4;10:340 [FREE Full text] [doi: 10.3389/fendo.2019.00340] [Medline: 31214119]
38. Crocetti E, Mattioli V, Buzzoni C, Franceschi S, Serraino D, Vaccarella S, for AIRTUM working group. Risk of thyroid as a first or second primary cancer. A population-based study in Italy, 1998-2012. Cancer Med 2021 Oct 1;10(19):6855-6867 [FREE Full text] [doi: 10.1002/cam4.4193] [Medline: 34533289]
39. Lal G, Groff M, Howe J, Weigel R, Sugg S, Lynch C. Risk of subsequent primary thyroid cancer after another malignancy: latency trends in a population-based study. Ann Surg Oncol 2012 Jun 1;19(6):1887-1896 [FREE Full text] [doi: 10.1245/s10434-011-2193-2] [Medline: 22227921]
40. Lian C, Zhuo R, Zhou R, Yu Y, Zhou P, Lin Q, et al. Risk factors of early thyroid dysfunction after definitive radiotherapy in nasopharyngeal carcinoma patients. Head Neck 2023 Sep 1;45(9):2344-2354 [FREE Full text] [doi: 10.1002/hed.27448] [Medline: 37415457]
41. Wang L, Jin Y, Zhou J, Pang C, Wang Y, Zhang S. Phenotypic disease network-based multimorbidity analysis in idiopathic cardiomyopathy patients with hospital discharge records. J Clin Med 2022 Nov 25;11(23):6965 [FREE Full text] [doi: 10.3390/jcm11236965] [Medline: 36498544]
42. Cruz-Ávila HA, Vallejo M, Martínez-García M, Hernández-Lemus E. Comorbidity networks in cardiovascular diseases. Front Physiol 2020 Aug 28;11:1009 [FREE Full text] [doi: 10.3389/fphys.2020.01009] [Medline: 32982776]
43. MIMIC-IV data sets. PhysioNet. URL: https://physionet.org/content/mimiciv/2.0/ [accessed 2024-09-23]


## Abbreviations

**BTD:** benign thyroid disease
**EMR:** electronic medical record
**ICD-10:** International Classification of Diseases, 10th Revision
**MIMIC:** Medical Information Mart for Intensive Care
**PCN:** phenotypic comorbidity network
**TC:** thyroid cancer
**TD:** thyroid disease

**TN:** thyroid nodule

*Edited by T de Azevedo Cardoso; submitted 27.11.23; peer-reviewed by H Qiu, B Zhang, R Levinson; comments to author 18.03.24; revised version received 16.05.24; accepted 10.09.24; published 03.10.24.*

*Please cite as:*
*Huang Y, Chen S, Wang Y, Ou X, Yan H, Gan X, Wei Z*
*Analyzing Comorbidity Patterns in Patients With Thyroid Disease Using Large-Scale Electronic Medical Records: Network-Based Retrospective Observational Study*
*Interact J Med Res 2024;13:e54891*
*URL: https://www.i-jmr.org/2024/1/e54891*
*doi:10.2196/54891*
*PMID:39361379*

Original Paper

# Factors Associated With Worsened Mental Health of Health Care Workers in Canada During the COVID-19 Pandemic: Cross-Sectional Survey Study

Ali AbdulHussein[1], BEng, MEng, MBA; Zahid Ahmad Butt[2], PhD; Stanko Dimitrov[1], PhD; Brian Cozzarin[1], PhD

[1]Department of Management Science, Faculty of Engineering, University of Waterloo, Waterloo, ON, Canada

[2]School 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@uwaterloo.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 as compared to HCWs 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 demographic 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.

***(Interact J Med Res 2024;13:e50064)*** 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.

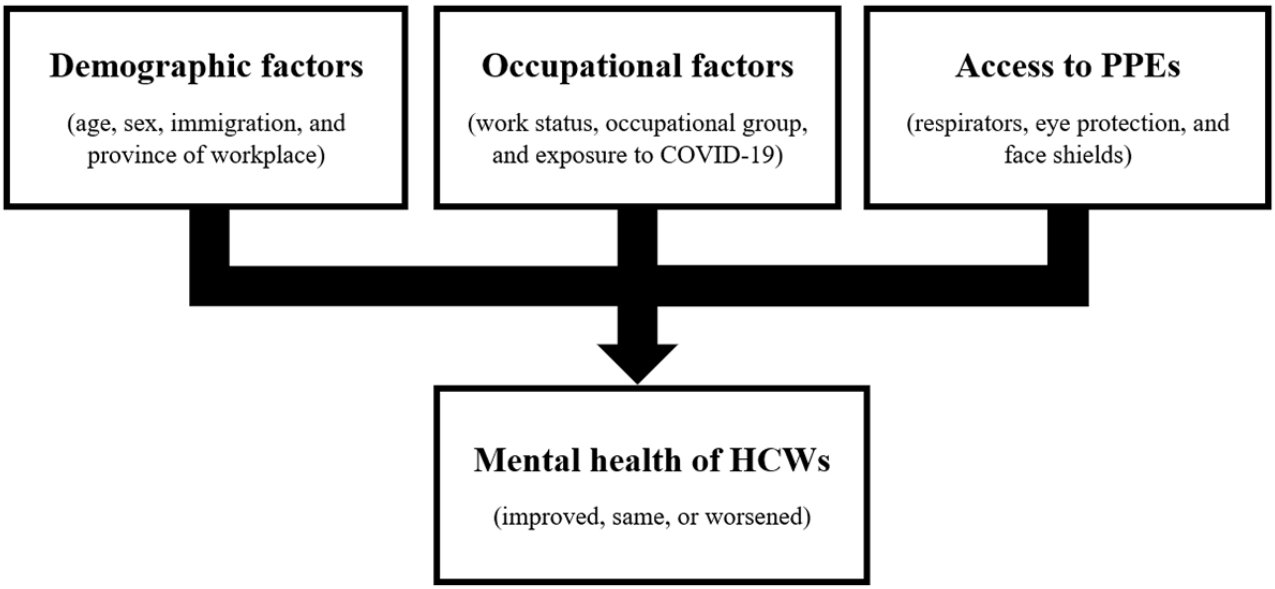
The questionnaire asked HCWs for information related to the job environment, demographics, geography, and information on access to PPE as background information. Adaptive questioning was used. Our study was limited to these 3 factors categories only: demographics, occupational, and access to PPE based 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[a] access characteristics.

| Characteristics | Value, n (%) |
|---|---|
| **Age (years)** | |
| <35 | 4261 (30.5) |
| 35-44 | 3962 (28.3) |
| 45-54 | 3415 (24.4) |
| ≥55 | 2352 (16.8) |
| **Sex** | |
| Male | 1308 (9.3) |
| Female | 12,682 (90.7) |
| **Immigration status** | |
| Nonimmigrant | 12,510 (89.4) |
| Immigrant | 1480 (10.6) |
| **Province of workplace** | |
| Atlantic provinces | 1972 (14.1) |
| Quebec | 832 (5.9) |
| Ontario | 6626 (47.4) |
| Manitoba and Saskatchewan | 1308 (9.3) |
| Alberta | 2018 (14.4) |
| British Columbia | 1194 (8.5) |
| Northern territories | 40 (0.3) |
| **Work status** | |
| Full-time | 10,152 (72.6) |
| Part-time | 3838 (27.4) |
| **Frontline occupation** | |
| Physician | 397 (2.8) |
| Nurse | 4689 (33.5) |
| Emergency medical personnel | 223 (1.6) |
| **Nonfrontline occupation** | |
| Personal support worker | 454 (3.2) |
| Allied health professional | 5511 (39.4) |
| Laboratory worker | 1267 (9.1) |
| Pharmacist | 169 (1.2) |
| Dental professional | 1280 (9.1) |
| **Exposure to confirmed or suspected cases** | |
| Yes | 5802 (41.5) |
| No | 8188 (58.5) |
| **Access to PPE** | |
| **Face shields** | |
| Always or usually available | 10,758 (76.9) |
| Sometimes available | 1070 (7.6) |
| Never available | 357 (2.6) |
| Skipped answer | 1805 (12.9) |
| **Eye protection** | |

| Characteristics | Value, n (%) |
|---|---|
| Always or usually available | 10,118 (72.3) |
| Sometimes available | 1132 (8.1) |
| Never available | 556 (4.0) |
| Skipped answer | 2184 (15.6) |
| **Respirators** | |
| Always or usually available | 6898 (49.3) |
| Sometimes available | 1369 (9.8) |
| Never available | 676 (4.8) |
| Skipped answer | 5047 (36.1) |

[a]PPE: personal protective equipment.

## 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[a].

| Independent variables | β | SE | OR[b] (95% CI) | *P* value |
|---|---|---|---|---|
| **Age group (years; reference: 35 to 44 years)** | | | | |
| Less than 35 | 0.13 | 0.05 | 1.14 (1.03-1.27) | .01 |
| 45-54 | –0.35 | 0.05 | 0.71 (0.64-0.78) | <.001 |
| 55 and older | –0.60 | 0.06 | 0.55 (0.49-0.61) | <.001 |
| **Sex (reference: female)** | | | | |
| Male | –0.28 | 0.06 | 0.76 (0.67-0.86) | <.001 |
| **Immigration status (reference: nonimmigrant)** | | | | |
| Immigrant | –0.56 | 0.06 | 0.57 (0.51-0.64) | <.001 |
| **Work location (reference: Alberta)** | | | | |
| Atlantic provinces | –0.41 | 0.07 | 0.66 (0.57-0.77) | <.001 |
| Quebec | –0.30 | 0.09 | 0.74 (0.61-0.89) | <.001 |
| Manitoba and Saskatchewan | –0.09 | 0.08 | 0.91 (0.77-1.08) | .27 |
| Ontario | –0.24 | 0.06 | 0.78 (0.70-0.88) | <.001 |
| British Columbia | –0.09 | 0.09 | 0.92 (0.77-1.09) | .31 |
| North Territories | –0.56 | 0.34 | 0.57 (0.30-1.15) | .01 |
| **Work status (reference: full-time)** | | | | |
| Part-time | –0.16 | 0.04 | 0.85 (0.76-0.93) | <.001 |
| **Job setting (reference: nonfrontline)** | | | | |
| Frontline | 0.23 | 0.05 | 1.26 (1.16-1.38) | <.001 |
| **Contact with patients with COVID-19 (reference: no contact)** | | | | |
| Exposure | 0.44 | 0.05 | 1.55 (1.41-1.70) | <.001 |
| **Access to respirators (reference: did not need)** | | | | |
| Always | –0.03 | 0.05 | 0.97 (0.88-1.07) | .13 |
| Sometimes | 0.30 | 0.08 | 1.34 (1.14-1.59) | <.001 |
| Never | 0.27 | 0.10 | 1.31 (1.07-1.62) | <.001 |
| **Access to eye protection (reference: did not need)** | | | | |
| Always | –0.03 | 0.06 | 0.97 (0.85-1.10) | .17 |
| Sometimes | 0.29 | 0.11 | 1.33 (1.08-1.65) | <.001 |
| Never | 0.41 | 0.13 | 1.51 (1.17-1.96) | <.001 |
| **Access to face protection (reference: did not need)** | | | | |
| Always | 0.26 | 0.07 | 1.30 (1.14-1.48) | <.001 |
| Sometimes | 0.40 | 0.11 | 1.50 (1.21-1.86) | <.001 |
| Never | 0.35 | 0.15 | 1.41 (1.05-1.92) | .02 |

[a]HCW: health care worker.

[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 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 [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.
[DOCX File , 21 KB - ijmr_v13i1e50064_app1.docx ]

### References

1. WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020. World Health Organization. 2020. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2023-03-01]
2. Overview: COVID-19's impact on health care systems. Canadian Institute for Health Information. 2021. URL: https://www.cihi.ca/en/covid-19-resources/impact-of-covid-19-on-canadas-health-care-systems/the-big-picture [accessed 2023-05-11]
3. Dudevich A, Frood J. Impact of the COVID-19 pandemic on health system use in Canada. Healthc Q 2021;24(2):12-14 [FREE Full text] [doi: 10.12927/hcq.2021.26552] [Medline: 34297658]
4. Carazo S, Pelletier M, Talbot D, Jauvin N, De Serres G, Vézina M. Psychological distress of healthcare workers in Québec (Canada) during the second and the third pandemic waves. J Occup Environ Med 2022;64(6):495-503 [FREE Full text] [doi: 10.1097/JOM.0000000000002487] [Medline: 35051960]
5. Brophy JT, Keith MM, Hurley M, McArthur JE. Sacrificed: Ontario healthcare workers in the time of COVID-19. New Solut 2021;30(4):267-281 [FREE Full text] [doi: 10.1177/1048291120974358] [Medline: 33174768]

6. Stuijfzand S, Deforges C, Sandoz V, Sajin CT, Jaques C, Elmers J, et al. Psychological impact of an epidemic/pandemic on the mental health of healthcare professionals: a rapid review. BMC Public Health 2020;20(1):1230 [FREE Full text] [doi: 10.1186/s12889-020-09322-z] [Medline: 32787815]
7. Vizheh M, Qorbani M, Arzaghi SM, Muhidin S, Javanmard Z, Esmaeili M. The mental health of healthcare workers in the COVID-19 pandemic: a systematic review. J Diabetes Metab Disord 2020;19(2):1967-1978 [FREE Full text] [doi: 10.1007/s40200-020-00643-9] [Medline: 33134211]
8. Liu JJW, Nazarov A, Plouffe RA, Forchuk CA, Deda E, Gargala D, et al. Exploring the well-being of health care workers during the COVID-19 pandemic: protocol for a prospective longitudinal study. JMIR Res Protoc 2021;10(9):e32663 [FREE Full text] [doi: 10.2196/32663] [Medline: 34477557]
9. Cai H, Tu B, Ma J, Chen L, Fu L, Jiang Y, et al. Psychological impact and coping strategies of frontline medical staff in hunan between january and march 2020 during the outbreak of coronavirus disease 2019 (COVID-19) in Hubei, China. Med Sci Monit 2020;26:e924171 [FREE Full text] [doi: 10.12659/MSM.924171] [Medline: 32291383]
10. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? Int J Epidemiol 2020;49(3):717-726 [FREE Full text] [doi: 10.1093/ije/dyaa033] [Medline: 32086938]
11. De Brier N, Stroobants S, Vandekerckhove P, De Buck E. Factors affecting mental health of health care workers during coronavirus disease outbreaks (SARS, MERS and COVID-19): a rapid systematic review. PLoS One 2020;15(12):e0244052 [FREE Full text] [doi: 10.1371/journal.pone.0244052] [Medline: 33320910]
12. Badahdah A, Khamis F, Al Mahyijari N, Al Balushi M, Al Hatmi H, Al Salmi I, et al. The mental health of health care workers in Oman during the COVID-19 pandemic. Int J Soc Psychiatry 2021;67(1):90-95 [FREE Full text] [doi: 10.1177/0020764020939596] [Medline: 32635837]
13. Yıldırım M, Arslan G, Özaslan A. Perceived risk and mental health problems among healthcare professionals during COVID-19 pandemic: exploring the mediating effects of resilience and coronavirus fear. Int J Ment Health Addict 2022;20(2):1035-1045 [FREE Full text] [doi: 10.1007/s11469-020-00424-8] [Medline: 33223977]
14. Serrano-Ripoll MJ, Meneses-Echavez JF, Ricci-Cabello I, Fraile-Navarro D, Fiol-deRoque MA, Pastor-Moreno G, et al. Impact of viral epidemic outbreaks on mental health of healthcare workers: a rapid systematic review and meta-analysis. J Affect Disord 2020;277:347-357 [FREE Full text] [doi: 10.1016/j.jad.2020.08.034] [Medline: 32861835]
15. Gupta B, Sharma V, Kumar N, Mahajan A. Anxiety and sleep disturbances among health care workers during the COVID-19 pandemic in India: cross-sectional online survey. JMIR Public Health Surveill 2020;6(4):e24206 [FREE Full text] [doi: 10.2196/24206] [Medline: 33284784]
16. Mrklas K, Shalaby R, Hrabok M, Gusnowski A, Vuong W, Surood S, et al. Prevalence of perceived stress, anxiety, depression, and obsessive-compulsive symptoms in health care workers and other workers in Alberta during the COVID-19 pandemic: cross-sectional survey. JMIR Ment Health 2020;7(9):e22408 [FREE Full text] [doi: 10.2196/22408] [Medline: 32915764]
17. Hadley J, Waidmann T. Health insurance and health at age 65: implications for medical care spending on new Medicare beneficiaries. Health Serv Res 2006;41(2):429-451 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00491.x] [Medline: 16584457]
18. Yunchao C, Yusof SA, Amin RM, Arshad MNM. Household debt and household spending behavior: evidence from Malaysia. J Ekon Malays 2020;54(1):111-120. [doi: 10.17576/jem-2020-5401-8]
19. Lancee WJ, Maunder RG, Goldbloom DS, Coauthors for the Impact of SARS Study. Prevalence of psychiatric disorders among Toronto hospital workers one to two years after the SARS outbreak. Psychiatr Serv 2008;59(1):91-95 [FREE Full text] [doi: 10.1176/ps.2008.59.1.91] [Medline: 18182545]
20. Maunder RG, Lancee WJ, Balderson KE, Bennett JP, Borgundvaag B, Evans S, et al. Long-term psychological and occupational effects of providing hospital healthcare during SARS outbreak. Emerg Infect Dis 2006;12(12):1924-1932 [FREE Full text] [doi: 10.3201/eid1212.060584] [Medline: 17326946]
21. Park J, Lee E, Park N, Choi YH. Mental health of nurses working at a government-designated hospital during a MERS-CoV outbreak: a cross-sectional study. Arch Psychiatr Nurs 2018;32(1):2-6 [FREE Full text] [doi: 10.1016/j.apnu.2017.09.006] [Medline: 29413067]
22. Huang JZ, Han MF, Luo TD, Ren AK, Zhou XP. [Mental health survey of medical staff in a tertiary infectious disease hospital for COVID-19]. Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi 2020;38(3):192-195 [FREE Full text] [doi: 10.3760/cma.j.cn121094-20200219-00063] [Medline: 32131151]
23. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
24. Zhang SX, Liu J, Jahanshahi AA, Nawaser K, Yousefi A, Li J, et al. At the height of the storm: healthcare staff's health conditions and job satisfaction and their associated predictors during the epidemic peak of COVID-19. Brain Behav Immun 2020;87:144-146 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.010] [Medline: 32387345]

25. Van Wert MJ, Gandhi S, Gupta I, Singh A, Eid SM, Haroon Burhanullah M, et al. Healthcare worker mental health after the initial peak of the COVID-19 pandemic: a US Medical Center cross-sectional survey. J Gen Intern Med 2022;37(5):1169-1176 [FREE Full text] [doi: 10.1007/s11606-021-07251-0] [Medline: 34993856]
26. Alshekaili M, Hassan W, Al Said N, Al Sulaimani F, Jayapal SK, Al-Mawali A, et al. Factors associated with mental health outcomes across healthcare settings in Oman during COVID-19: frontline versus non-frontline healthcare workers. BMJ Open 2020;10(10):e042030 [FREE Full text] [doi: 10.1136/bmjopen-2020-042030] [Medline: 33040019]
27. Cai Q, Feng H, Huang J, Wang M, Wang Q, Lu X, et al. The mental health of frontline and non-frontline medical workers during the coronavirus disease 2019 (COVID-19) outbreak in China: a case-control study. J Affect Disord 2020;275:210-215 [FREE Full text] [doi: 10.1016/j.jad.2020.06.031] [Medline: 32734910]
28. Abdessater M, Rouprêt M, Misrai V, Matillon X, Gondran-Tellier B, Freton L, et al. COVID19 pandemic impacts on anxiety of French urologist in training: outcomes from a national survey. Prog Urol 2020;30(8-9):448-455 [FREE Full text] [doi: 10.1016/j.purol.2020.04.015] [Medline: 32376208]
29. Lu W, Wang H, Lin Y, Li L. Psychological status of medical workforce during the COVID-19 pandemic: a cross-sectional study. Psychiatry Res 2020;288:112936 [FREE Full text] [doi: 10.1016/j.psychres.2020.112936] [Medline: 32276196]
30. Labrague LJ, de Los Santos JAA. Fear of COVID-19, psychological distress, work satisfaction and turnover intention among frontline nurses. J Nurs Manag 2021;29(3):395-403 [FREE Full text] [doi: 10.1111/jonm.13168] [Medline: 32985046]
31. Fournier A, Laurent A, Lheureux F, Ribeiro-Marthoud MA, Ecarnot F, Binquet C, et al. Impact of the COVID-19 pandemic on the mental health of professionals in 77 hospitals in France. PLoS One 2022;17(2):e0263666 [FREE Full text] [doi: 10.1371/journal.pone.0263666] [Medline: 35171915]
32. Marco CA, Larkin GL, Feeser VR, Monti JE, Vearrier L, ACEP Ethics Committee. Post-traumatic stress and stress disorders during the COVID-19 pandemic: survey of emergency physicians. J Am Coll Emerg Physicians Open 2020;1(6):1594-1601 [FREE Full text] [doi: 10.1002/emp2.12305] [Medline: 33392568]
33. Oudyk J, Smith P. Occupational Health Clinics for Ontario workers mayday, mayday workplace mental health series 2020. Occupational Health Clinics for Ontario. 2020. URL: https://www.ohcow.on.ca/mayday-mayday-2020.html [accessed 2023-06-05]
34. Greenberg N. Mental health of health-care workers in the COVID-19 era. Nat Rev Nephrol 2020;16(8):425-426 [FREE Full text] [doi: 10.1038/s41581-020-0314-5] [Medline: 32561871]
35. Lewis M, Palmer VJ, Kotevski A, Densley K, O'Donnell ML, Johnson C, et al. Rapid design and delivery of an experience-based co-designed mobile app to support the mental health needs of health care workers affected by the COVID-19 pandemic: impact evaluation protocol. JMIR Res Protoc 2021;10(3):e26168 [FREE Full text] [doi: 10.2196/26168] [Medline: 33635823]
36. Vizheh M, Qorbani M, Arzaghi SM, Muhidin S, Javanmard Z, Esmaeili M. The mental health of healthcare workers in the COVID-19 pandemic: a systematic review. J Diabetes Metab Disord 2020;19(2):1967-1978 [FREE Full text] [doi: 10.1007/s40200-020-00643-9] [Medline: 33134211]
37. David E, DePierro JM, Marin DB, Sharma V, Charney DS, Katz CL. COVID-19 pandemic support programs for healthcare workers and implications for occupational mental health: a narrative review. Psychiatr Q 2022;93(1):227-247 [FREE Full text] [doi: 10.1007/s11126-021-09952-5] [Medline: 34606067]
38. Siskind D, Harris M, Pirkis J, Whiteford H. Personalised support delivered by support workers for people with severe and persistent mental illness: a systematic review of patient outcomes. Epidemiol Psychiatr Sci 2012;21(1):97-110. [doi: 10.1017/s2045796011000734] [Medline: 22670417]
39. Deif R, Salama M. Depression from a precision mental health perspective: utilizing personalized conceptualizations to guide personalized treatments. Front Psychiatry 2021;12:650318 [FREE Full text] [doi: 10.3389/fpsyt.2021.650318] [Medline: 34045980]
40. Fiol-DeRoque MA, Serrano-Ripoll MJ, Jiménez R, Zamanillo-Campos R, Yáñez-Juan AM, Bennasar-Veny M, et al. A mobile phone-based intervention to reduce mental health problems in health care workers during the COVID-19 pandemic (PsyCovidApp): randomized controlled trial. JMIR Mhealth Uhealth 2021;9(5):e27039 [FREE Full text] [doi: 10.2196/27039] [Medline: 33909587]


## Abbreviations

**HCW:** health care worker
**OR:** odds ratio
**PPE:** personal protective equipment

*Edited by T de Azevedo Cardoso; submitted 18.06.23; peer-reviewed by S Alwan; comments to author 08.11.23; revised version received 26.11.23; accepted 12.01.24; published 15.02.24.*

*Please cite as:*
*AbdulHussein A, Butt ZA, Dimitrov S, Cozzarin B*
*Factors Associated With Worsened Mental Health of Health Care Workers in Canada During the COVID-19 Pandemic: Cross-Sectional Survey Study*
*Interact J Med Res 2024;13:e50064*
*URL: https://www.i-jmr.org/2024/1/e50064*
*doi:10.2196/50064*
*PMID:38358785*

Original Paper

# A Web-Based Tool to Assess Social Inclusion and Support Care Planning in Mental Health Supported Accommodation: Development and Preliminary Test Study

Sharon Eager[1], MSc; Helen Killaspy[1,2], MBBS, PhD; Joanna C[1], MSc; Gillian Mezey[3], MBBS; Peter McPherson[1], DPsych; Megan Downey[2], BSc; Georgina Thompson[1], MSc; Brynmor Lloyd-Evans[1], PhD

[1]Division of Psychiatry, University College London, London, United Kingdom

[2]Camden and Islington NHS Foundation Trust, London, United Kingdom

[3]Population Health Research Institute, St George's, University of London, London, United Kingdom

**Corresponding Author:**
Brynmor Lloyd-Evans, PhD
Division of Psychiatry
University College London
Maple House
149 Tottenham Court Road
London, W1T 7NF
United Kingdom
Phone: 44 (0)20 7679 9428
Email: b.lloyd-evans@ucl.ac.uk

## 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.

*(Interact J Med Res 2024;13:e45987)* doi:10.2196/45987

**KEYWORDS**

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 based 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 a 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 initial impressions 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 semistructured 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 1.** Laboratory testing and field testing of the online Social Inclusion Questionnaire User Experience—characteristics of participants.

| Participant characteristics | Laboratory testing | Field testing | |
|---|---|---|---|
| | Staff (n=6), n (%) | Staff (n=12), n (%) | Service users (n=12), n (%) |
| **Gender** | | | |
| Male | 2 (33) | 5 (42) | 11 (92) |
| Female | 4 (67) | 6 (50) | 0 (0) |
| Nonbinary | 0 (0) | 1 (8) | 1 (8) |
| **Age group (y)** | | | |
| 18-30 | 0 (0) | 8 (67) | 4 (33) |
| 31-50 | 2 (33) | 2 (17) | 3 (25) |
| ≥51 | 4 (67) | 1 (8) | 4 (33) |
| Prefer not to say | 0 (0) | 1 (8) | 1 (8) |
| **Ethnicity** | | | |
| Asian/Asian British | 1 (17) | 0 (0) | 1 (8) |
| Black/Black British | 3 (50) | 2 (17) | 4 (33) |
| White/White British | 2 (33) | 8 (67) | 1 (8) |
| Mixed/multiple ethnic groups | 0 (0) | 1 (8) | 6 (50) |
| Other ethnic background | 0 (0) | 1 (8) | 0 (0) |
| **Sexual orientation** | | | |
| Heterosexual | N/A[a] | N/A | 6 (50) |
| Gay/lesbian | N/A | N/A | 1 (8) |
| Bisexual | N/A | N/A | 2 (17) |
| Prefer not to say | N/A | N/A | 3 (25) |
| **Type of supported accommodation lived or worked in** | | | |
| Floating outreach support | 2 (33) | 2 (17) | 2 (17) |
| 9-to-5 supported housing | 2 (33) | 1 (8) | 2 (17) |
| 24-h supported housing | 2 (33) | 7 (58) | 8 (67) |
| Residential care | 0 (0) | 2 (17) | 0 (0) |
| **Length of time worked or lived in supported accommodation (y)** | | | |
| <2 | 1 (17) | 7 (58) | 5 (42) |
| 2-5 | 3 (50) | 3 (25) | 5 (42) |
| 6-10 | 1 (17) | 2 (17) | 0 (0) |
| ≥10 | 1 (17) | 0 (0) | 1 (8) |
| Prefer not to say | 0 (0) | 0 (0) | 1 (8) |

[a]N/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.

| Section of the SInQUE affected | Explanation of the problem | Resolution | Justification for the change |
|---|---|---|---|
| **Registration and use changes** | | | |
| The online SInQUE home page | Some staff members and service users suggested developing additional materials to help explain the SInQUE[a]. | 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. | It is easier for managers, commissioners, staff, and service users to understand and use the SInQUE. |
| The initial page where staff members are asked to register for the SInQUE | 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. | Changed response options to a fixed-response drop-down menu with all housing providers in the borough and an "other" free-text option. | Allows for compilation of service-level data and is easier for staff to navigate. |
| The page where staff members enter details to set up a new SInQUE assessment | 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. | Additional guidance on how the questionnaire should be administered was added to the introduction paragraph of the SInQUE. | Important contextual information for the questionnaire was explicitly clarified. |
| The page where staff members enter details to set up a new SInQUE assessment | 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. | Changed the wording of the question to the following: "Please select the type of housing support the service user receives from the list below." | The clarity of the question improved. |
| **Changes to the wording of SInQUE questions** | | | |
| The section covering "leisure" questions | 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?" | Changed the wording of this subquestion to the following: "Other leisure activity?" | The clarity of the question improved. |
| The section covering "leisure" questions | 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]. | 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?" | The clarity and appropriateness of the question improved. |
| The section covering "social" questions | 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. | Changed the wording of the question to the following: "How many people, outside the workers in your care team, could you confide in?" | The clarity of the question improved. |
| The section covering "home life/housing" questions | Some staff members thought that question 36—"What kind of accommodation do you live in?"—was worded ambiguously. | Changed the wording of the question to the following: "What type of housing support do you receive?" | The clarity of the question improved. |
| The section covering "home life/housing" questions | 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. | Changed statement to the following: "Question omitted, not included in the online SInQUE." | The clarity of the question improved. |
| **Changes to the SInQUE summary outputs** | | | |
| The SInQUE summary report | 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. | Simplified the spider graph to show percentage of total score in the graph instead of frequency. | It was easier to interpret the graph as sections with different totals became standardized. |
| The SInQUE summary report for multiple assessments completed with the same service user | In the section summarizing scores across multiple time points, staff members thought that a visual depiction of this comparison would be useful. | Made comparative bar charts of multiple scores across time with the same service user available on the summary report. | It was easier to understand and relay the results. |

[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:

- Improve staff relationship with and understanding of the service user
- Help plan more relevant, targeted support for the service user

- Long term:
  - Borough-level improvements and changes in services to support social inclusion
  - Individual-level benefits for service users' recovery and social inclusion

### *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 SInQUE. 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 supported accommodation 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).

| Activity | Implementation goal being addressed | COM-B[a] domain |
|---|---|---|
| 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. | Increase awareness of the SInQUE tool among supported accommodation staff and respond to any of their concerns or other problems. | • Motivation<br>• Opportunity<br>• Capability |
| 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. | 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 | • Motivation |
| 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. | Lack of supported accommodation management prioritization for staff to use the SInQUE | • Motivation<br>• Opportunity |
| Local authority commissioners contact all supported accommodation managers to encourage the use of the SInQUE within services in the borough. | Lack of supported accommodation management prioritization for staff to use the SInQUE | • Motivation |
| 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. | Increase visibility and awareness of the SInQUE among supported accommodation managers and respond to any of their concerns or other problems. | • Motivation<br>• Opportunity |
| Study research assistant (SE) offers supported accommodation staff technological support with SInQUE registration and use. | Uncertainty about how to manage the technical process of using the SInQUE | • Capability |
| 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. | Uncertainty among some supported accommodation staff members about how best to explain the purpose of the SInQUE and engage service users | • Capability |
| 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]. | 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. | • Motivation |
| 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]. | 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. | • Motivation |

[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.

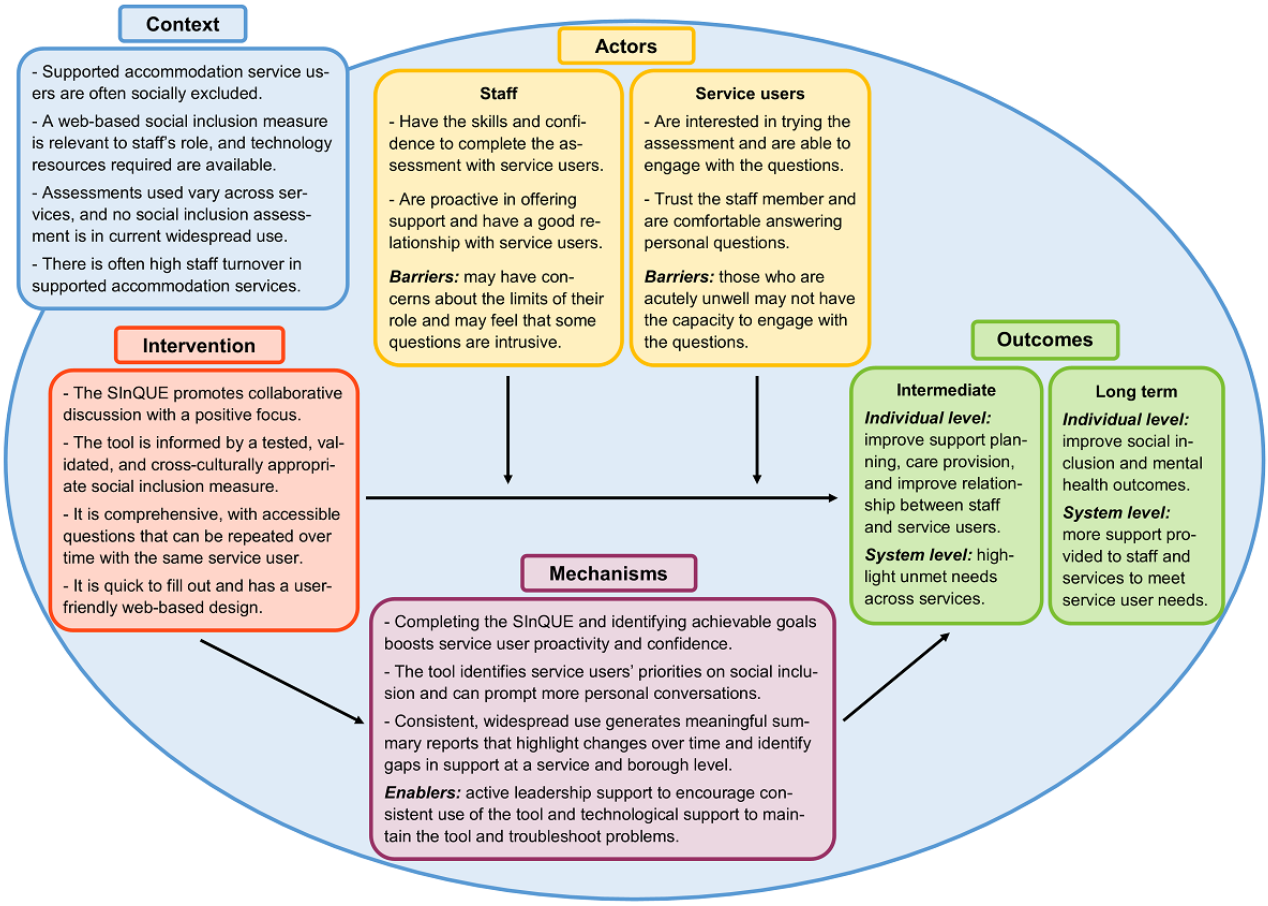
## Usage Data

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).

**Figure 1.** Intervention-context-actor-mechanism-outcome logic model for the online Social Inclusion Questionnaire User Experience (SInQUE).




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

means of implementation support. 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.

[DOCX File , 22 KB - ijmr_v13i1e45987_app5.docx ]

## References


1. Burchardt T, Le Grand J, Piachaud D. Social exclusion in Britain 1991—1995. Soc Policy Adm 2002 Dec 17;33(3):227-244. [doi: 10.1111/1467-9515.00148]
2. Le Boutillier C, Croucher A. Social inclusion and mental health. Brit J Occup Ther 2010 Mar 01;73(3):136-139. [doi: 10.4276/030802210x12682330090578]
3. Morgan C, Burns T, Fitzpatrick R, Pinfold V, Priebe S. Social exclusion and mental health: conceptual and methodological review. Br J Psychiatry 2007 Dec;191:477-483. [doi: 10.1192/bjp.bp.106.034942] [Medline: 18055950]
4. Wright N, Stickley T. Concepts of social inclusion, exclusion and mental health: a review of the international literature. J Psychiatr Ment Health Nurs 2013 Feb;20(1):71-81. [doi: 10.1111/j.1365-2850.2012.01889.x] [Medline: 22369652]
5. Huxley P, Thornicroft G. Social inclusion, social quality and mental illness. Br J Psychiatry 2003 Apr;182:289-290. [doi: 10.1192/bjp.182.4.289] [Medline: 12668401]
6. Koenders JF, de Mooij LD, Dekker JM, Kikkert M. Social inclusion and relationship satisfaction of patients with a severe mental illness. Int J Soc Psychiatry 2017 Dec;63(8):773-781. [doi: 10.1177/0020764017737572] [Medline: 29067839]
7. McManus S, Meltzer H, Brugha T, Bebbington PE, Jenkins R. Adult psychiatric morbidity in England: results of a household survey. Health and Social Care Information Centre. 2009. URL: https://openaccess.city.ac.uk/id/eprint/23647/1/adul-psyc-morb-res-hou-sur-eng-2007-rep.pdf [accessed 2022-12-07]
8. Luciano A, Meara E. Employment status of people with mental illness: national survey data from 2009 and 2010. Psychiatr Serv 2014 Oct;65(10):1201-1209 [FREE Full text] [doi: 10.1176/appi.ps.201300335] [Medline: 24933361]
9. Boardman J, Grove B, Perkins R, Shepherd G. Work and employment for people with psychiatric disabilities. Br J Psychiatry 2003 Jun;182:467-468. [doi: 10.1192/bjp.182.6.467] [Medline: 12777333]
10. de Mooij LD, Kikkert M, Lommerse NM, Peen J, Meijwaard SC, Theunissen J, et al. Victimisation in adults with severe mental illness: prevalence and risk factors. Br J Psychiatry 2015 Dec;207(6):515-522. [doi: 10.1192/bjp.bp.113.143370] [Medline: 26450584]
11. Kamperman AM, Henrichs J, Bogaerts S, Lesaffre EM, Wierdsma AI, Ghauharali RR, et al. Criminal victimisation in people with severe mental illness: a multi-site prevalence and incidence survey in the Netherlands. PLoS One 2014;9(3):e91029 [FREE Full text] [doi: 10.1371/journal.pone.0091029] [Medline: 24609108]
12. Boardman J. Social exclusion and mental health – how people with mental health problems are disadvantaged: an overview. Ment Health Soc Incl 2011;15(3):112-121 [FREE Full text] [doi: 10.1108/20428301111165690]
13. Mezey G, White S, Harrison I, Bousfield J, Killaspy H, Lloyd-Evans B, et al. 'Modelling social exclusion in a diagnostically-mixed sample of people with severe mental illness'. Int J Soc Psychiatry 2022 Mar;68(2):420-428 [FREE Full text] [doi: 10.1177/00207640211001893] [Medline: 33730906]
14. Ware NC, Hopper K, Tugenberg T, Dickey B, Fisher D. Connectedness and citizenship: redefining social integration. Psychiatr Serv 2007 Apr;58(4):469-474. [doi: 10.1176/ps.2007.58.4.469] [Medline: 17412847]
15. Guidance for commissioners of rehabilitation services for people with complex mental health needs. Joint Commissioning Panel for Mental Health. 2016 Nov. URL: https://www.rcpsych.ac.uk/docs/default-source/members/faculties/rehabilitation-and-social-psychiatry/rehab-social—joint-commissioning-panel—guidance-for-commissioners-of-rehabilitation-services-for-people-with-complex-mental-health-needs—2016.pdf?sfvrsn=82cbb5e4_6 [accessed 2022-11-03]
16. Killaspy H, Priebe S. Research into mental health supported accommodation - desperately needed but challenging to deliver. Br J Psychiatry 2021 Apr;218(4):179-181. [doi: 10.1192/bjp.2020.74] [Medline: 32321596]
17. Killaspy H. Supported accommodation for people with mental health problems. World Psychiatry 2016 Feb;15(1):74-75 [FREE Full text] [doi: 10.1002/wps.20278] [Medline: 26833612]
18. Krotofil J, McPherson P, Killaspy H. Service user experiences of specialist mental health supported accommodation: a systematic review of qualitative studies and narrative synthesis. Health Soc Care Community 2018 Nov;26(6):787-800. [doi: 10.1111/hsc.12570] [Medline: 29609195]
19. Bitter NA, Roeg DP, van Nieuwenhuizen C, van Weeghel J. Identifying profiles of service users in housing services and exploring their quality of life and care needs. BMC Psychiatry 2016 Nov 23;16(1):419 [FREE Full text] [doi: 10.1186/s12888-016-1122-0] [Medline: 27881159]
20. de Heer-Wunderink C, Visser E, Caro-Nienhuis A, Sytema S, Wiersma D. Supported housing and supported independent living in the Netherlands, with a comparison with England. Community Ment Health J 2012 Jun;48(3):321-327 [FREE Full text] [doi: 10.1007/s10597-011-9381-1] [Medline: 21246274]
21. McPherson P, Krotofil J, Killaspy H. Mental health supported accommodation services: a systematic review of mental health and psychosocial outcomes. BMC Psychiatry 2018 May 15;18(1):128 [FREE Full text] [doi: 10.1186/s12888-018-1725-8] [Medline: 29764420]
22. Bitter N, Roeg D, van Nieuwenhuizen C, van Weeghel J. Recovery in supported accommodations: a scoping review and synthesis of interventions for people with severe mental illness. Community Ment Health J 2020 Aug;56(6):1053-1076 [FREE Full text] [doi: 10.1007/s10597-020-00561-3] [Medline: 32016620]

23. Piat M, Sofouli E, Sabetti J, Lambrou A, Chodos H, Briand C, et al. Protocol for a mixed studies systematic review on the implementation of the recovery approach in adult mental health services. BMJ Open 2017 Aug 30;7(8):e017080 [FREE Full text] [doi: 10.1136/bmjopen-2017-017080] [Medline: 28855202]
24. Macpherson R, Shepherd G, Thyarappa P. Supported accommodation for people with severe mental illness: an update. Adv Psychiatr Treat 2012;18(5):381-391. [doi: 10.1192/apt.bp.110.008714]
25. Sandhu S, Priebe S, Leavey G, Harrison I, Krotofil J, McPherson P, et al. Intentions and experiences of effective practice in mental health specific supported accommodation services: a qualitative interview study. BMC Health Serv Res 2017 Jul 11;17(1):471 [FREE Full text] [doi: 10.1186/s12913-017-2411-0] [Medline: 28693490]
26. Killaspy H, Priebe S, McPherson P, Zenasni Z, Greenberg L, McCrone P, et al. Predictors of moving on from mental health supported accommodation in England: national cohort study. Br J Psychiatry 2020 Jun;216(6):331-337. [doi: 10.1192/bjp.2019.101] [Medline: 31046864]
27. Roe D, Slade M, Jones N. The utility of patient-reported outcome measures in mental health. World Psychiatry 2022 Feb;21(1):56-57 [FREE Full text] [doi: 10.1002/wps.20924] [Medline: 35015343]
28. Guidance and technical packages on community mental health services: promoting person-centred and rights-based approaches. World Health Organization. 2021 May 26. URL: https://www.who.int/publications/i/item/guidance-and-technical-packages-on-community-mental-health-services [accessed 2022-09-11]
29. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
30. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
31. Killaspy H, White S, Dowling S, Krotofil J, McPherson P, Sandhu S, et al. Adaptation of the Quality Indicator for Rehabilitative Care (QuIRC) for use in mental health supported accommodation services (QuIRC-SA). BMC Psychiatry 2016 Apr 14;16:101 [FREE Full text] [doi: 10.1186/s12888-016-0799-4] [Medline: 27075574]
32. Mezey G, White S, Thachil A, Berg R, Kallumparam S, Nasiruddin O, et al. Development and preliminary validation of a measure of social inclusion for use in people with mental health problems: the SInQUE. Int J Soc Psychiatry 2013 Aug;59(5):501-507. [doi: 10.1177/0020764012443752] [Medline: 22511274]
33. Baumgartner JN, Burns JK. Measuring social inclusion--a key outcome in global mental health. Int J Epidemiol 2014 Apr;43(2):354-364. [doi: 10.1093/ije/dyt224] [Medline: 24306926]
34. Mezey G, White S, Harrison I, Bousfield J, Lloyd-Evans B, Payne S, et al. Validity, reliability, acceptability, and utility of the Social Inclusion Questionnaire User Experience (SInQUE): a clinical tool to facilitate social inclusion amongst people with severe mental health problems. Soc Psychiatry Psychiatr Epidemiol 2020 Jul;55(7):953-964 [FREE Full text] [doi: 10.1007/s00127-019-01826-3] [Medline: 32016512]
35. SInQUE home page. SInQUE. URL: https://sinque.org.uk [accessed 2022-12-07]
36. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348:g1687 [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
37. Steare T, O'Hanlon P, Eskinazi M, Osborn D, Lloyd-Evans B, Jones R, et al. Smartphone-delivered self-management for first-episode psychosis: the ARIES feasibility randomised controlled trial. BMJ Open 2020 Aug 26;10(8):e034927 [FREE Full text] [doi: 10.1136/bmjopen-2019-034927] [Medline: 32847902]
38. Mukumbang FC, Marchal B, Van Belle S, van Wyk B. Using the realist interview approach to maintain theoretical awareness in realist studies. Qual Res 2020;20(4):485-515. [doi: 10.1177/1468794119881985]
39. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
40. Priebe S, Saidi M, Want A, Mangalore R, Knapp M. Housing services for people with mental disorders in England: patient characteristics, care provision and costs. Soc Psychiatry Psychiatr Epidemiol 2009 Oct;44(10):805-814. [doi: 10.1007/s00127-009-0001-0] [Medline: 19277440]
41. Pinfold V, Sweet D, Porter I, Quinn C, Byng R, Griffiths C, et al. Improving community health networks for people with severe mental illness: a case study investigation. National Institute for Health Research. 2015. URL: https://eprints.whiterose.ac.uk/84845/2/ScientificSummary_hsdr03050.pdf [accessed 2022-07-12]


## Abbreviations

**ICAMO:** intervention-context-actor-mechanism-outcome
**NHS:** National Health Service
**REC:** Research Ethics Committee
**SInQUE:** Social Inclusion Questionnaire User Experience
**TIDieR:** Template for Intervention Description and Replication
**UCL:** University College London

*Edited by T de Azevedo Cardoso; submitted 25.01.23; peer-reviewed by P Dabas, G Leavey, L Yang, B Hannigan; comments to author 14.11.23; revised version received 08.12.23; accepted 10.02.24; published 13.03.24.*

*Please cite as:*
*Eager S, Killaspy H, C J, Mezey G, McPherson P, Downey M, Thompson G, Lloyd-Evans B*
*A Web-Based Tool to Assess Social Inclusion and Support Care Planning in Mental Health Supported Accommodation: Development and Preliminary Test Study*
*Interact J Med Res 2024;13:e45987*
*URL: https://www.i-jmr.org/2024/1/e45987*
*doi:10.2196/45987*
*PMID:38477978*

Original Paper

# Patterns of Skin Picking in Skin Picking Disorder: Ecological Momentary Assessment Study

Christina Gallinat[1], PhD; Markus Moessner[1], PD, PhD; Maximilian Wilhelm[1,2], MSc; Nancy Keuthen[3], PhD; Stephanie Bauer[1,2,4], PD, PhD

[1]Center for Psychotherapy Research, University Hospital Heidelberg, Heidelberg, Germany

[2]German Center for Mental Health (DZPG), partner site Mannheim-Heidelberg-Ulm, Heidelberg, Germany

[3]Trichotillomania and Excoriation Disorder Program, Center for OCD and Related Disorders, Massachusetts General Hospital, Boston, MA, United States

[4]Heidelberg 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: $n_{episodes}$=642, 43.8%; 20 min: $n_{episodes}$=312, 21.3%; 30 min: $n_{episodes}$=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: $n_{episodes}$=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 [$M_f$]=4.01, SD 0.69 vs mean automatic style [$M_a$]=3.47, SD 0.99; $P$<.001; SMD=0.64; desire to pick: $M_f$=2.61, SD 1.06 vs $M_a$=1.94, SD 1.03; $P$<.001; SMD=0.82), while boredom and concentration problems were more prominent in automatic skin picking (boredom: $M_f$=1.69, SD 0.89 vs $M_a$=1.84, SD 0.89; $P$=.03; SMD=–0.31; concentration problems: $M_f$=2.06, SD 0.87 vs $M_a$=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

***(Interact J Med Res 2024;13:e53831)*** doi:10.2196/53831

**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 almost 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 intraindividual and interindividual 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 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 ($\alpha$=0.84; subscales: symptom severity: $\alpha$=0.77 and impairment: $\alpha$=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 ($\alpha$=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 $\alpha$=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 ($\alpha$=0.84).

### Anxiety

Symptoms of generalized anxiety disorders were assessed with the Generalized Anxiety Disorders-7 (GAD-7) [26]. The Cronbach α 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 (eg, 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).

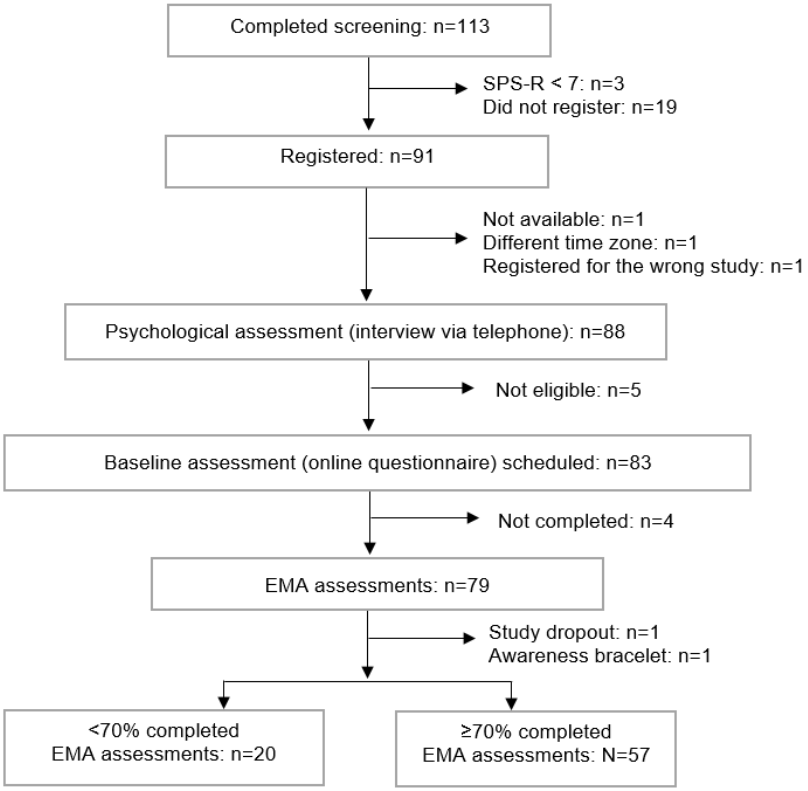
## Results

### 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 (ie, 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).

**Figure 1.** Participant flow. EMA: ecological momentary assessment; SPS-R: Skin Picking Scale-Revised.




## 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 $\chi$-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.

| Characteristics | Total EMA[a] sample (n=77) | EMA≥70% (n=57) | EMA<70% (n=20) |
|---|---|---|---|
| Female sex, n (%) | 74 (96.1) | 54 (94.7) | 20 (100) |
| Age (years), mean (SD) | 28.84 (6.51) | 29.3 (6.77) | 27.3 (5.74) |
| **Education, n (%)** | | | |
| Still in school | 1 (1.3) | —[b] | 1 (5) |
| Middle secondary | 8 (10.4) | 6 (10.5) | 2 (10) |
| Highest secondary | 27 (35.1) | 19 (33.3) | 8 (40) |
| University | 41 (53.2) | 32 (56.1) | 9 (45) |
| **Occupational status, n (%)** | | | |
| Employed | 40 (51.9) | 28 (49.1) | 12 (60) |
| Trainee | 1 (1.3) | 1 (1.8) | — |
| School student | 1 (1.3) | — | 1 (5) |
| University student | 23 (29.9) | 18 (31.6) | 5 (25) |
| Housewife or househusband | 3 (3.9) | 3 (5.3) | — |
| Retired | 2 (2.6) | 1 (1.8) | 1 (5) |
| Unemployed | 2 (2.6) | 1 (1.8) | 1 (5) |
| Other | 5 (6.5) | 5 (8.8) | — |
| **Family status** | | | |
| Single, n (%) | 34 (44.2) | 22 (38.6) | 12 (60) |
| In a relationship, n (%) | 25 (32.5) | 20 (35.1) | 5 (25) |
| Married, n (%) | 15 (19.5) | 12 (21.1) | 3 (15) |
| Separated or divorced, n (%) | 2 (2.6) | 2 (3.5) | — |
| Other, n (%) | 1 (1.3) | 1 (1.8) | — |
| PHQ-9[c], mean (SD) | 11.95 (5.6) | 11.63 (5.41) | 12.15 (5.86) |
| GAD-7[d], mean (SD) | 10.04 (4.64) | 9.63 (4.85) | 10.5 (4.01) |
| SPS-R[e], mean (SD) | 17.69 (3.98) | 18 (4) | 16.7 (3.87) |
| SPIS[f], mean (SD) | 23.45 (8.57) | 23.33 (8.94) | 23.8 (7.61) |
| MIDAS focused[g], mean (SD) | 19.9 (3.92) | 19.7 (3.9) | 20.45 (4.05) |
| MIDAS automatic[h], mean (SD) | 18.18 (3.63) | 18.49 (3.58) | 17.3 (3.74) |

[a]EMA: ecological momentary assessment.

[b]Not available.

[c]PHQ-9: Patient Health Questionnaire-9, depressive symptoms.

[d]GAD-7: generalized anxiety disorders-7.

[e]SPS-R: Skin Picking Scale-Revised.

[f]SPIS: Skin Picking Impact Scale.

[g]MIDAS focused: Milwaukee Inventory for the Dimensions of Adult Skin Picking, focused skin picking.

[h]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; $n_{episodes}$=253, 17.3% of all 1467 episodes) and the last (8-10 PM; $n_{episodes}$=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.

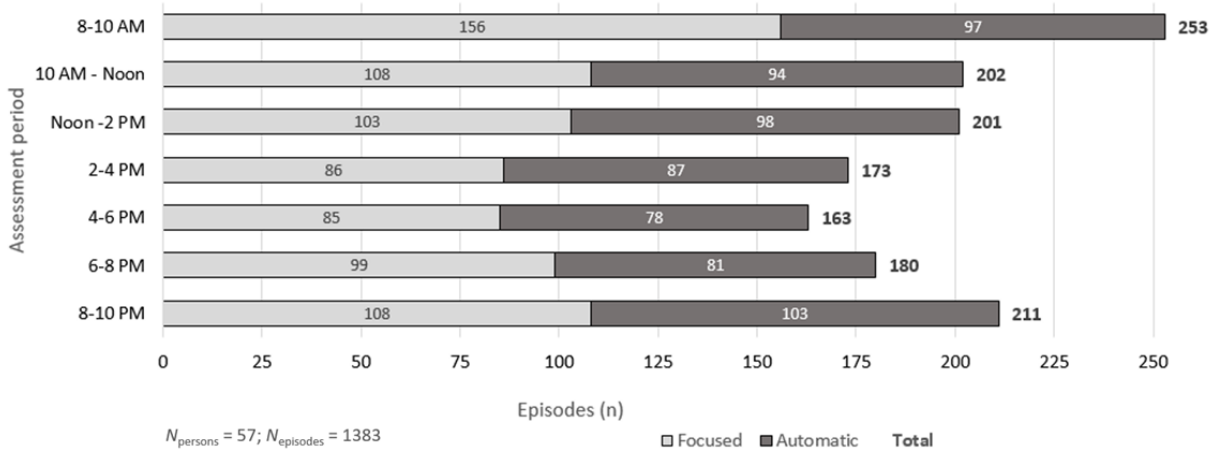



**Table 2.** Urge and episode parameters over the day[a].

| Time | All episodes (N=1467), n (%) | Focused episodes (n=806), n (%) | Automatic episodes (n=661), n (%) | Intensity of episodes (n=1467), mean (SD) | Urge intensity EMA[b] with episodes (n=1467), mean (SD) | Urge intensity EMA without episodes (n=2291), mean (SD) |
|---|---|---|---|---|---|---|
| 8-10 AM | 253 (17.2) | 156 (19.35) | 97 (14.67) | 2.61 (0.87) | 2.53 (0.98) | 1.78 (0.79) |
| 10 AM to Noon | 202 (13.8) | 108 (13.4) | 94 (14.22) | 2.38 (0.72) | 2.54 (0.92) | 1.92 (0.73) |
| Noon to 2 PM | 201 (13.7) | 103 (12.78) | 98 (14.83) | 2.19 (0.78) | 2.75 (0.86) | 1.87 (0.68) |
| 2-4 PM | 173 (11.8) | 86 (10.67) | 87 (13.16) | 2.43 (0.86) | 2.89 (0.84) | 1.83 (0.71) |
| 4-6 PM | 163 (11.1) | 85 (10.55) | 78 (11.8) | 2.43 (0.88) | 2.90 (0.95) | 1.92 (0.76) |
| 6-8 PM | 180 (12.3) | 99 (12.28) | 81 (12.25) | 2.43 (0.78) | 3 (0.85) | 1.92 (0.71) |
| 8-10 PM | 211 (14.4) | 108 (13.4) | 103 (15.58) | 2.77 (0.74) | 3.02 (0.88) | 1.94 (0.76) |

[a]Only periods of time-based assessments are listed. The total N refers to all registered episodes ($N_{episodes}$=1467, $N_{persons}$=57). Intensity scores are average scores of person-wise means.

[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[a].

| Approximate length | Episodes), n (%) |
|---|---|
| 10 min | 642 (43.8) |
| 20 min | 312 (21.3) |
| 30 min | 217 (14.8) |
| 40 min | 97 (6.6) |
| 50 min | 39 (2.7) |
| 60 min | 64 (4.4) |
| 70 min | 14 (0.9) |
| 80 min | 26 (1.8) |
| 90 min | 22 (1.4) |
| 100 min | 6 (0.4) |
| 110 min | 2 (0.1) |
| 120 min | 26 (1.8) |

[a]All episodes reported in time- and event-based ecological momentary assessments ($N_{episodes}$=1467; $N_{persons}$=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; $t_{56}$=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 [$M_f$]=2.56, SD 0.62; mean automatic style [$M_a$]=2.45, SD 0.78; $t_{48}$=1.52, $P$=.14; urge intensity: $M_f$=2.86, SD 0.77; $M_a$=2.90, SD 0.86; $t_{48}$=–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.

**Table 4.** Self-reported triggers[a].

| Trigger | Total (N=57), mean (SD) | Focused (n=49), mean (SD) | Automatic (n=49), mean (SD) | *t* test[b] (*df*) | *P* value | SMD |
|---|---|---|---|---|---|---|
| Visual or tactile cues | 3.64 (1.26) | 4.01 (0.69) | 3.47 (0.99) | 4.482 (48) | <.001 | 0.64 |
| Tension | 2.63 (1.29) | 2.67 (0.89) | 2.84 (0.97) | –1.532 (48) | .13 | –0.22 |
| Boredom | 1.69 (1.03) | 1.69 (0.89) | 1.84 (0.89) | –2.187 (48) | .03 | –0.31 |
| Problems with concentration | 2.17 (1.31) | 2.06 (0.87) | 2.31(1.06) | –2.847 (48) | .006 | –0.41 |
| Wanted to pick the skin | 2.14 (1.27) | 2.61 (1.06) | 1.94 (1.03) | 5.753 (48) | <.001 | 0.82 |
| Habit or routine | 2.71 (1.45) | 2.82 (1.07) | 2.56 (1.10) | 1.818 (48) | .08 | 0.26 |
| Itch | 1.67 (1.09) | 1.75 (0.94) | 1.81 (1) | –0.594 (39) | .556 | –0.09 |

[a]Answers 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 ($n_{episodes}$=1295, $n_{persons}$=49). Further, 8 persons were excluded from the comparison as they reported no automatic episodes. "Habit/routine" relates to the item "I picked my skin out of a routine (eg, after arriving home or during the evening bath routine)."

[b]2-tailed.

## Discussion

### 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 a few 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% ($n_{episodes}$=642/1467) were no more than 10 minutes long and 80% ($n_{episodes}$=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, ≥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

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, is it 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 abovementioned 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


1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5 ed. Washington: American Psychiatric Publishing; 2013.
2. Grant JE, Chamberlain SR. Prevalence of skin picking (excoriation) disorder. J Psychiatr Res 2020;130:57-60 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.06.033] [Medline: 32781374]
3. Tucker BT, Woods DW, Flessner CA, Franklin SA, Franklin ME. The skin picking impact project: phenomenology, interference, and treatment utilization of pathological skin picking in a population-based sample. J Anxiety Disord 2011;25(1):88-95. [doi: 10.1016/j.janxdis.2010.08.007] [Medline: 20810239]
4. Flessner CA, Woods DW. Phenomenological characteristics, social problems, and the economic impact associated with chronic skin picking. Behav Modif 2006;30(6):944-963. [doi: 10.1177/0145445506294083] [Medline: 17050772]
5. Gallinat C, Stürmlinger LL, Schaber S, Bauer S. Pathological skin picking: phenomenology and associations with emotions, self-esteem, body image, and subjective physical well-being. Front Psychiatry 2021;12:732717 [FREE Full text] [doi: 10.3389/fpsyt.2021.732717] [Medline: 34721107]
6. Wilhelm S, Keuthen NJ, Deckersbach T, Engelhard IM, Forker AE, Baer L, et al. Self-injurious skin picking: clinical characteristics and comorbidity. J Clin Psychiatry 1999;60(7):454-459. [Medline: 10453800]
7. Walther MR, Flessner CA, Conelea CA, Woods DW. The Milwaukee inventory for the dimensions of adult skin picking (MIDAS): initial development and psychometric properties. J Behav Ther Exp Psychiatry 2009;40(1):127-135. [doi: 10.1016/j.jbtep.2008.07.002] [Medline: 18725154]
8. Lin A, Farhat LC, Flores JM, Levine JL, Fernandez TV, Bloch MH, et al. Characteristics of trichotillomania and excoriation disorder across the lifespan. Psychiatry Res 2023;322:115120. [doi: 10.1016/j.psychres.2023.115120] [Medline: 36842397]
9. Neziroglu F, Rabinowitz D, Breytman A, Jacofsky M. Skin picking phenomenology and severity comparison. Prim Care Companion 2008;10(4):306-312 [FREE Full text] [doi: 10.4088/pcc.v10n0406] [Medline: 18787665]
10. Odlaug BL, Grant JE. Clinical characteristics and medical complications of pathologic skin picking. Gen Hosp Psychiatry 2008;30(1):61-66. [doi: 10.1016/j.genhosppsych.2007.07.009] [Medline: 18164942]
11. Prochwicz K, Kałużna-Wielobób A, Kłosowska J. Skin picking in a non-clinical sample of young Polish adults. Prevalence and characteristics. Compr Psychiatry 2016;71:77-85. [doi: 10.1016/j.comppsych.2016.08.013] [Medline: 27639125]
12. Bohne A, Wilhelm S, Keuthen NJ, Baer L, Jenike MA. Skin picking in German students: prevalence, phenomenology, and associated characteristics. Behav Modif 2002;26(3):320-339. [doi: 10.1177/0145445502026003002] [Medline: 12080904]
13. Ebner-Priemer UW, Trull TJ. Ambulatory assessment: an innovative and promising approach for clinical psychology. Eur Psychol 2009;14(2):109-119. [doi: 10.1027/1016-9040.14.2.109]
14. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
15. Walz LC, Nauta MH, aan het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: a systematic review. J Anxiety Disord 2014;28(8):925-937. [doi: 10.1016/j.janxdis.2014.09.022] [Medline: 25445083]
16. Serre F, Fatseas M, Swendsen J, Auriacombe M. Ecological momentary assessment in the investigation of craving and substance use in daily life: a systematic review. Drug Alcohol Depend 2015;148:1-20. [doi: 10.1016/j.drugalcdep.2014.12.024] [Medline: 25637078]
17. Engel SG, Crosby RD, Thomas G, Bond D, Lavender JM, Mason T, et al. Ecological momentary assessment in eating disorder and obesity research: a review of the recent literature. Curr Psychiatry Rep 2016;18(4):37. [doi: 10.1007/s11920-016-0672-7] [Medline: 26893235]
18. Wilhelm M, Feldhege J, Bauer S, Moessner M. Use of internet-based monitoring in psychotherapy research [Einsatz internetbasierter Verlaufsmessung in der Psychotherapieforschung.]. Psychother 2020;65(6):505-511. [doi: 10.1007/s00278-020-00461-7]
19. Snorrason I, Ólafsson RP, Flessner CA, Keuthen NJ, Franklin ME, Woods DW. The skin picking scale-revised: factor structure and psychometric properties. J Obsessive-Compulsive Relat Disord 2012;1(2):133-137. [doi: 10.1016/j.jocrd.2012.03.001]
20. Gallinat C, Keuthen N, Backenstrass M. A self-report instrument for the assessment of dermatillomania: reliability and validity of the German skin picking scale-revised [Ein Selbstbeurteilungsinstrument zur Erfassung von Dermatillomanie: Reliabilität und Validität der deutschsprachigen Version der Skin Picking Scale-Revised]. Psychother Psychosom Med Psychol 2016;66(6):249-255. [doi: 10.1055/s-0042-107255] [Medline: 27286530]
21. Margraf J, Cwik JC, von Brachel R, Suppiger A, Schneider S. DIPS Open Access 1.2: Diagnostic Interview for Mental Disorders [DIPS Open Access 1.2: Diagnostisches Interview bei psychischen Störungen]. Bochum: Mental Health Research and Treament Center, Ruhr-Universität Bochum; 2021. URL: https://omp.ub.rub.de/index.php/RUB/catalog/view/172/149/972 [accessed 2024-07-02]

22. Keuthen NJ, Deckersbach T, Wilhelm S, Engelhard I, Forker A, O'Sullivan RL, et al. The Skin Picking Impact Scale (SPIS): scale development and psychometric analyses. Psychosomatics 2001;42(5):397-403. [doi: 10.1176/appi.psy.42.5.397] [Medline: 11739906]
23. Mehrmann LM, Hunger A, Gerlach AL. Excoriation disorder (skin picking): an initial analysis of disorder-specific assessment instruments. [Pathologisches Hautzupfen/ -quetschen (skin picking): Erste Ergebnisse zur Psychometrie störungsspezifischer Messinstrumente] [Article in German]. Klin Psychol Psychother 2017;46(1):23-31. [doi: 10.1026/1616-3443/a000386]
24. Gallinat C, Moessner M, Haenssle HA, Winkler JK, Backenstrass M, Bauer S. SaveMySkin: an internet-based self-help intervention for skin picking. Study protocol for a randomized pilot study. Contemp Clin Trials Commun 2019;13:100315 [FREE Full text] [doi: 10.1016/j.conctc.2018.100315] [Medline: 31011654]
25. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. JAMA 1999;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
26. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
27. Grant JE, Peris TS, Ricketts EJ, Lochner C, Stein DJ, Stochl J, et al. Identifying subtypes of trichotillomania (hair pulling disorder) and excoriation (skin picking) disorder using mixture modeling in a multicenter sample. J Psychiatr Res 2021;137:603-612 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.11.001] [Medline: 33172654]
28. Pozza A, Albert U, Dèttore D. Early maladaptive schemas as common and specific predictors of skin picking subtypes. BMC Psychol 2020;8(1):27 [FREE Full text] [doi: 10.1186/s40359-020-0392-y] [Medline: 32188511]
29. Grant JE, Bethlehem RAI, Chamberlain SR, Peris TS, Ricketts EJ, O'Neill J, et al. Neurobiology of subtypes of trichotillomania and skin picking disorder. CNS Spectr 2023;28(1):98-103 [FREE Full text] [doi: 10.1017/S109285292100095X] [Medline: 34730081]
30. Arnold LM, McElroy SL, Mutasim DF, Dwight MM, Lamerson CL, Morris EM. Characteristics of 34 adults with psychogenic excoriation. J Clin Psychiatry 1998;59(10):509-514. [doi: 10.4088/jcp.v59n1003] [Medline: 9818631]


## 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
**$M_a$:** mean automatic style (MIDAS)
**$M_f$:** 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


*Edited by T de Azevedo Cardoso; submitted 23.10.23; peer-reviewed by E Olfson, A Wabnegger; comments to author 22.11.23; revised version received 12.12.23; accepted 03.05.24; published 18.07.24.*

*Please cite as:*
*Gallinat C, Moessner M, Wilhelm M, Keuthen N, Bauer S*
*Patterns of Skin Picking in Skin Picking Disorder: Ecological Momentary Assessment Study*
*Interact J Med Res 2024;13:e53831*
*URL: https://www.i-jmr.org/2024/1/e53831*
*doi:10.2196/53831*
*PMID:*

Original Paper

# Effects of Background Music on Attentional Networks of Children With and Without Attention Deficit/Hyperactivity Disorder: Case Control Experimental Study

Camila Guimarães Mendes[1], PhD; Jonas Jardim de Paula[2], PhD; Débora Marques Miranda[1], PhD

[1]Department of Pediatrics, Federal University of Minas Gerais, Belo Horizonte, Brazil

[2]Department 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 ($F_{1,72=}9.83$; $P=.03$; $\eta p^2=0.06$) but the condition×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

***(Interact J Med Res 2024;13:e53869)*** doi:10.2196/53869

**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]. The

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

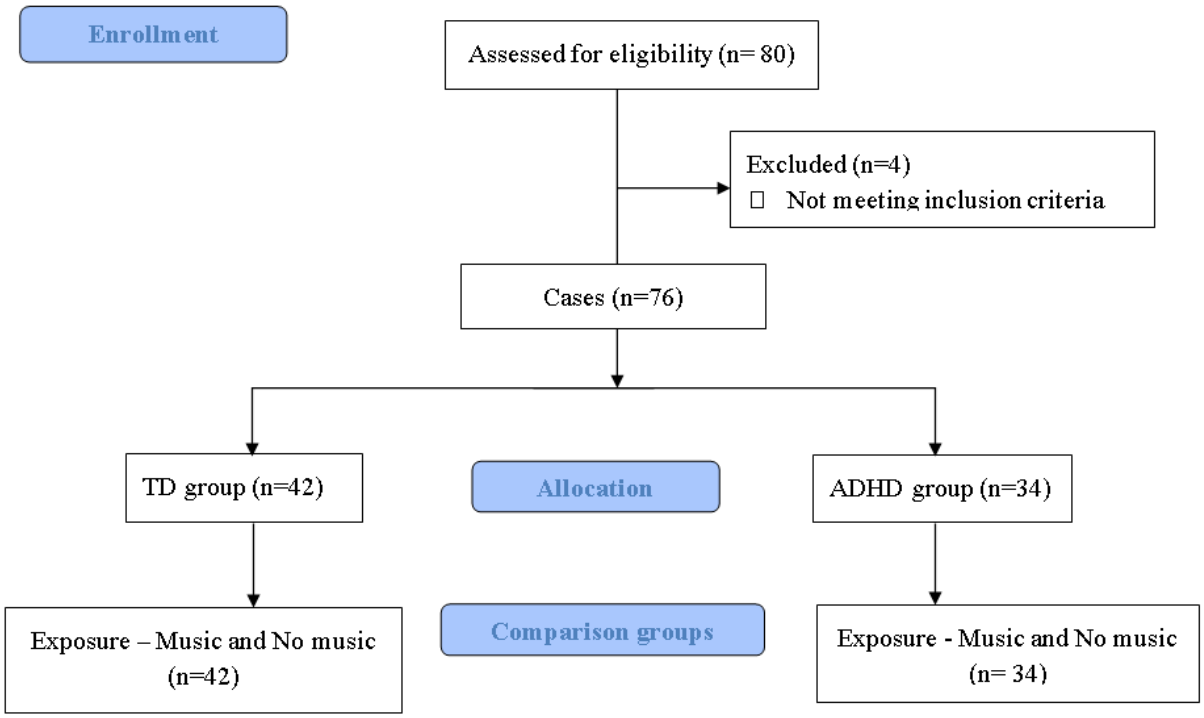
### Study Design

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.

| Order | Title | Duration (min:s) |
|---|---|---|
| 1 | Fortnite OST–Battle Royale Menu Music (Rock Version) [31][a] | 3:50 |
| 2 | Alone (Mashmallo) – Modified [32][b] | 3:19 |
| 3 | Free Fire New EPIC Theme Song [33][c] | 3:56 |
| 4 | Herobrine's Life (Instrumental) [34][d] | 4:00 |
| 5 | Olha a explosão (Mc Kevinho) – Modified [35][b] | 3:07 |

[a]This 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.

[b]The original song was modified with Audacity (version 2.3.2; Audacity Team), an audio editor, to remove the voices.

[c]This song is the theme song of Free Fire 2019.

[d]This 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.

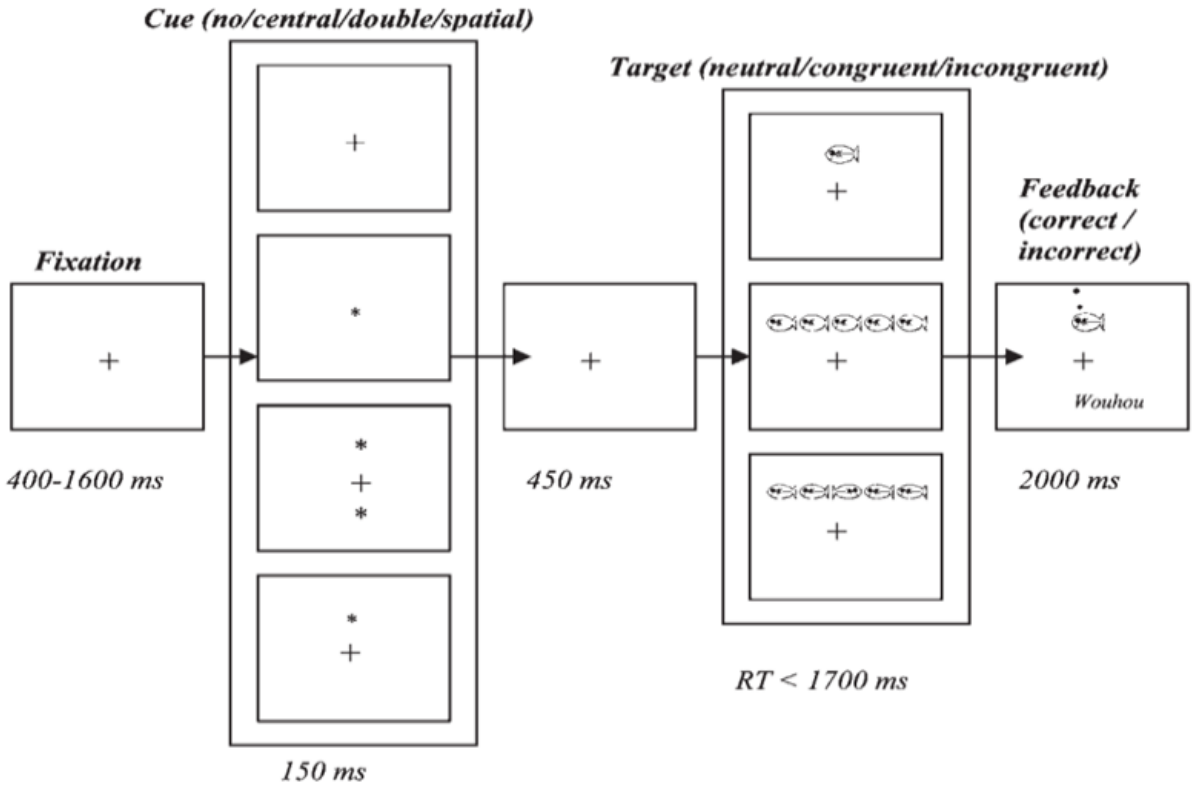
Additionally, all children and caregivers were asked to complete a semistructured questionnaire, which can be found in Multimedia Appendix 2, before starting the task.

## 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 3.010; Universität Düsseldorf), considering $\alpha$=.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].

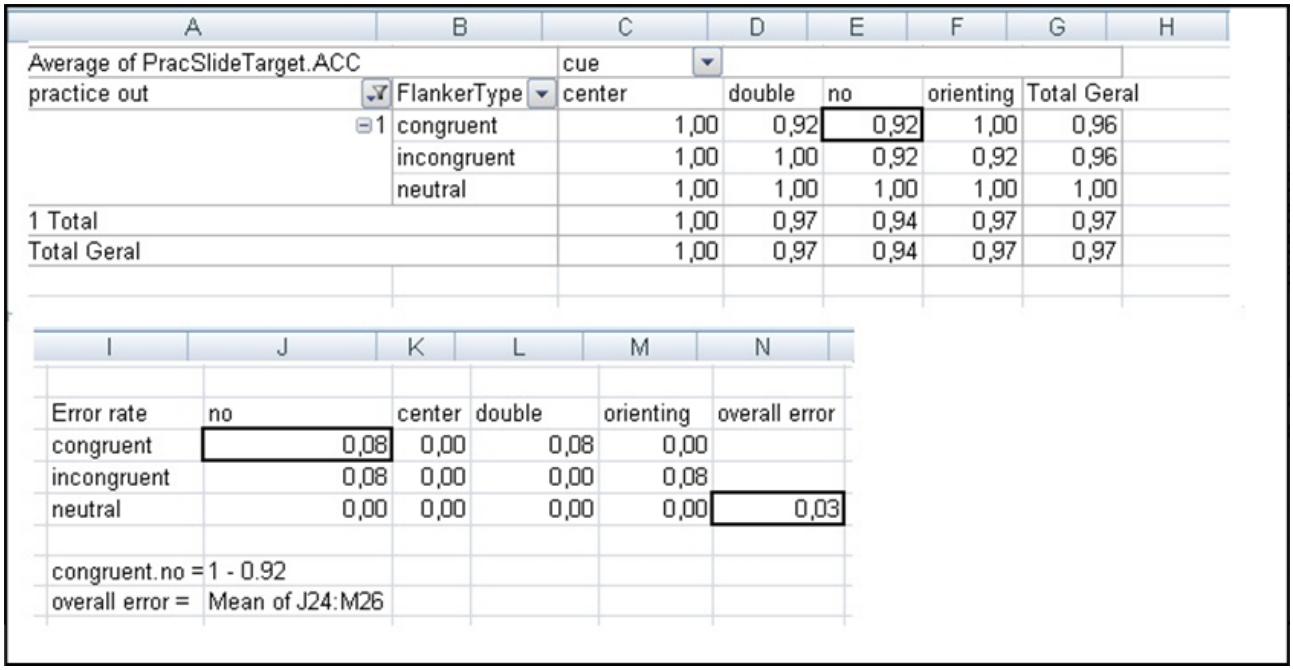
**Figure 4.** Example of calculations of attentional networks for participant 1, session 2, using an Excel (Microsoft Corp) macro.

| Average of PracSlideTarget.RT | | | | cue | | | | |
|---|---|---|---|---|---|---|---|---|
| outiler | practice | PracSli | FlankerType | center | double | no | orienting | Total Geral |
| 1 | 1 | 1 | congruent | 578 | 682 | 644 | 666 | 641 |
| | | | incongruent | 611 | 740 | 882 | 670 | 724 |
| | | | neutral | 592 | 633 | 573 | 653 | 613 |
| | | 1 Total | | 594 | 685 | 696 | 663 | 659 |
| | 1 Total | | | 594 | 685 | 696 | 663 | 659 |
| 1 Total | | | | 594 | 685 | 696 | 663 | 659 |
| Total Geral | | | | 594 | 685 | 696 | 663 | 659 |

| subject: | 1 | session: | 2 | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | no | center | double | orienting | | Alerting | **11** |
| | congruent | 644 | 578 | 682 | 666 | | Orienting | **-69** |
| | incongruen | 882 | 611 | 740 | 670 | | Conflict | **82** |
| | neutral | 573 | 592 | 633 | 653 | 660 | | |

**Alerting (696 - 685) = 11**
**Orienting (594 - 663) = -69**
**Conflict (724 - 641) = 82**

**Figure 5.** Example of calculations of overall errors of participant 1, session 2, using an Excel (Microsoft Corp) macro. The outlined squares indicate the values used for the calculations in the example.



| Average of PracSlideTarget.ACC | | cue | | | | |
|---|---|---|---|---|---|---|
| practice out | FlankerType | center | double | no | orienting | Total Geral |
| 1 | congruent | 1,00 | 0,92 | 0,92 | 1,00 | 0,96 |
| | incongruent | 1,00 | 1,00 | 0,92 | 0,92 | 0,96 |
| | neutral | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |
| 1 Total | | 1,00 | 0,97 | 0,94 | 0,97 | 0,97 |
| Total Geral | | 1,00 | 0,97 | 0,94 | 0,97 | 0,97 |

| Error rate | no | center | double | orienting | overall error |
|---|---|---|---|---|---|
| congruent | 0,08 | 0,00 | 0,08 | 0,00 | |
| incongruent | 0,08 | 0,00 | 0,00 | 0,08 | |
| neutral | 0,00 | 0,00 | 0,00 | 0,00 | 0,03 |
| congruent.no = | 1 - 0.92 | | | | |
| overall error = | Mean of J24:M26 | | | | |

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 ($t_{71}$=–1.158; *P*=.25).

**Table 2.** Demographic characteristics of participants (N=76).

| Characteristics | TD group (n=42), mean (SD) | ADHD group (n=34), mean (SD) |
|---|---|---|
| Age (years) | 11.0 (0.85) | 10.9 (0.75) |
| Brazilian Criteria of Economic Classification[a] | 28.5 (11.32) | 31.5 (11.15) |
| **SNAP-IV[b] symptoms (*t* score[c])** | | |
| Inattention | 1 (1.97) | 6 (2.35)) |
| Hyperactivity/impulsivity | 0.5 (0.84) | 5 (1.85) |
| CBCL-ADHD[d] (*t* score[c]) | 44.7 (5.90) | 61.6 (11.47) |
| **CCPT[e] score** | | |
| Omissions | 11.7 (9.5) | 11.9 (8.6) |
| Commissions | 25.0 (6.0) | 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; $\eta p^2$=0.06) but not for the group×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).

| Group and flanker condition | Reaction time with music (ms), mean (SD) | | | | Reaction time without music (ms), mean (SD) | | | |
|---|---|---|---|---|---|---|---|---|
| | Central cue | Double cue | No cue | Orienting | Central cue | Double cue | No cue | Orienting |
| **Typically developing group** | | | | | | | | |
| Congruent | 639 (120) | 626 (266) | 692 (130) | 619 (192) | 635 (135) | 614 (177) | 675 (229) | 612 (296) |
| Incongruent | 702 (106) | 689 (343) | 736 (201) | 680 (62) | 674 (75) | 680 (177) | 723 (109) | 660 (203) |
| Neutral | 621 (167) | 597 (115) | 664 (227) | 590 (182) | 634 (205) | 612 (255) | 657 (63) | 598 (246) |
| **Attention deficit/hyperactivity disorder group** | | | | | | | | |
| Congruent | 686 (120) | 668 (266) | 743 (130) | 663 (192) | 689 (120) | 672 (266) | 734 (130) | 648 (192) |
| Incongruent | 763 (106) | 722 (343) | 796 (201) | 693 (62) | 759 (106) | 708 (343) | 781 (201) | 710 (62) |
| Neutral | 672 (167) | 658 (115) | 732 (227) | 647 (182) | 651 (167) | 648 (115) | 709 (227) | 629 (182) |

**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).

| Measure | TD group score, mean (SD) | | ADHD group, mean (SD) | | Main effect of music | | Interaction of music×group | |
|---|---|---|---|---|---|---|---|---|
| | No music | Music | No music | Music | *F* test (*df*) | *P* value | *F* test (*df*) | *P* value |
| Alerting | 68.62 (52.67) | 60.33 (81.28) | 79.56 (54.64) | 62.32 (64.75) | 1.20 (1) | .28 | 0.15 (1) | .70 |
| Orienting | 26.69 (47.6) | 24.29 (40.88) | 38.62 (50.33) | 44.03 (57.8) | 0.86 (1) | .36 | 0.28 (1) | .60 |
| Conflict | 58.02 (47.06) | 53.81 (38.28) | 56.18 (47.35) | 56.50 (38.28) | 0.11 (1) | .74 | 0.12 (1) | .73 |
| Error rate[a] | 0.041 (0.036) | 0.039 (0.049) | 0.066 (0.058) | 0.052 (0.042) | 8.83 (1) | .03 | 1.79 (1) | .18 |

[a]$\eta p^2$=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.

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

1. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Washington, DC: American Psychiatric Association; 2013.
2. Martella D, Aldunate N, Fuentes LJ, Sánchez-Pérez N. Arousal and executive alterations in attention deficit hyperactivity disorder (ADHD). Front Psychol 2020;11:1991 [FREE Full text] [doi: 10.3389/fpsyg.2020.01991] [Medline: 32903419]
3. Wong KP, Qin J. Effectiveness of social virtual reality training in enhancing social interaction skills in children with attention-deficit/hyperactivity disorder: protocol for a three-arm pilot randomized controlled trial. JMIR Res Protoc 2023 Sep 18;12:e48208 [FREE Full text] [doi: 10.2196/48208] [Medline: 37721790]
4. Brown TE. ADD/ADHD and impaired executive function in clinical practice. Curr Psychiatry Rep 2008 Oct;10(5):407-411. [doi: 10.1007/s11920-008-0065-7] [Medline: 18803914]

5. Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. Biol Psychiatry 2005 Jun 01;57(11):1336-1346. [doi: 10.1016/j.biopsych.2005.02.006] [Medline: 15950006]
6. Amaravathi T, Murugan M, Subramanian A, Geetha A. Attention deficit hyperactivity disorder. Trends in Biosciences 2019;12(3):214-221. [doi: 10.4324/9780203065563-16]
7. Berger A, Posner MI. Pathologies of brain attentional networks. Neurosci Biobehav Rev 2000 Jan;24(1):3-5. [doi: 10.1016/s0149-7634(99)00046-9] [Medline: 10654653]
8. Sergeant JA. Modeling attention-deficit/hyperactivity disorder: a critical appraisal of the cognitive-energetic model. Biol Psychiatry 2005 Jun 01;57(11):1248-1255. [doi: 10.1016/j.biopsych.2004.09.010] [Medline: 15949995]
9. Zajenkowski M, Goryńska E, Winiewski M. Variability of the relationship between personality and mood. Pers Individ Dif 2012 May;52(7):858-861. [doi: 10.1016/j.paid.2012.01.007]
10. Martin-Moratinos M, Bella-Fernández M, Blasco-Fontecilla H. Effects of music on attention-deficit/hyperactivity disorder (ADHD) and potential application in serious video games: systematic review. J Med Internet Res 2023 May 12;25:e37742 [FREE Full text] [doi: 10.2196/37742] [Medline: 37171837]
11. Söderlund G, Sikström S, Smart A. Listen to the noise: noise is beneficial for cognitive performance in ADHD. J Child Psychol Psychiatry 2007 Aug;48(8):840-847. [doi: 10.1111/j.1469-7610.2007.01749.x] [Medline: 17683456]
12. Mendes CG, Diniz LA, Marques Miranda D. Does music listening affect attention? a literature review. Dev Neuropsychol 2021;46(3):192-212. [doi: 10.1080/87565641.2021.1905816] [Medline: 33813988]
13. Dalton BH, Behm DG. Effects of noise and music on human and task performance: A systematic review. Occup Ergon 2008 Jan 11;7(3):143-152. [doi: 10.3233/oer-2007-7301]
14. Husain G, Thompson W, Schellenberg E. Effects of musical tempo and mode on arousal, mood, and spatial abilities. Music Percept 2002;20(2):151-171. [doi: 10.1525/mp.2002.20.2.151]
15. Putkinen V, Makkonen T, Eerola T. Music-induced positive mood broadens the scope of auditory attention. Soc Cogn Affect Neurosci 2017 Jul 01;12(7):1159-1168 [FREE Full text] [doi: 10.1093/scan/nsx038] [Medline: 28460035]
16. Schellenberg E, Hallam S. Music listening and cognitive abilities in 10- and 11-year-olds: the blur effect. Ann N Y Acad Sci 2005 Dec;1060:202-209. [doi: 10.1196/annals.1360.013] [Medline: 16597767]
17. Gonzalez MF, Aiello JR. More than meets the ear: Investigating how music affects cognitive task performance. J Exp Psychol Appl 2019 Sep;25(3):431-444. [doi: 10.1037/xap0000202] [Medline: 30688499]
18. Hides L, Dingle G, Quinn C, Stoyanov SR, Zelenko O, Tjondronegoro D, et al. Efficacy and outcomes of a music-based emotion regulation mobile app in distressed young people: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11482 [FREE Full text] [doi: 10.2196/11482] [Medline: 30664457]
19. Abikoff H, Courtney ME, Szeibel PJ, Koplewicz HS. The effects of auditory stimulation on the arithmetic performance of children with ADHD and nondisabled children. J Learn Disabil 1996 May;29(3):238-246. [doi: 10.1177/002221949602900302] [Medline: 8732885]
20. Greenop K, Kann L. Extra-task stimulation on mathematics performance in children with and without ADHD. South African J Psychol 2016 Aug 02;37(2):330-344. [doi: 10.1177/008124630703700208]
21. Pelham WE, Waschbusch DA, Hoza B, Gnagy EM, Greiner AR, Sams SE, et al. Music and video as distractors for boys with ADHD in the classroom: comparison with controls, individual differences, and medication effects. J Abnorm Child Psychol 2011 Nov;39(8):1085-1098. [doi: 10.1007/s10802-011-9529-z] [Medline: 21695447]
22. Martin-Moratinos M, Bella-Fernández M, Blasco-Fontecilla H. Effects of music on attention-deficit/hyperactivity disorder (ADHD) and potential application in serious video games: systematic review. J Med Internet Res 2023 May 12;25:e37742 [FREE Full text] [doi: 10.2196/37742] [Medline: 37171837]
23. Fabio R, Urso M. The analysis of Attention Network in ADHD, attention problems and typically developing subjects. Life Span and Disability XVII 2014;17(2):199-221.
24. Adólfsdóttir S, Sørensen L, Lundervold AJ. The attention network test: a characteristic pattern of deficits in children with ADHD. Behav Brain Funct 2008 Feb 12;4(1):9 [FREE Full text] [doi: 10.1186/1744-9081-4-9] [Medline: 18269768]
25. Egermann H. Can parents influence children's music preferences and positively shape their development? Egermann.net. 2013. URL: http://www.egermann.net/wp-content/uploads/2018/11/Review_Egermann_Parents_Influence.pdf [accessed 2024-03-23]
26. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 1997 Jul;36(7):980-988 [FREE Full text] [doi: 10.1097/00004583-199707000-00021] [Medline: 9204677]
27. Pasquali L, Wechsler S, Bensusan E. Matrizes progressivas do Raven Infantil: um estudo de validação para o Brasil. Aval Psicol 2002;1(2):95-110 [FREE Full text]
28. Bordin IA, Rocha MM, Paula CS, Teixeira MCTV, Achenbach TM, Rescorla LA, et al. Child Behavior Checklist (CBCL), Youth Self-Report (YSR) and Teacher's Report Form (TRF): an overview of the development of the original and Brazilian versions. Cad Saude Publica 2013 Jan;29(1):13-28 [FREE Full text] [doi: 10.1590/s0102-311x2013000100004] [Medline: 23370021]

29. Costa DS, de Paula JJ, Malloy-Diniz LF, Romano-Silva MA, Miranda DM. Parent SNAP-IV rating of attention-deficit/hyperactivity disorder: accuracy in a clinical sample of ADHD, validity, and reliability in a Brazilian sample. J Pediatr (Rio J) 2019;95(6):736-743 [FREE Full text] [doi: 10.1016/j.jped.2018.06.014] [Medline: 30236592]
30. Bradley MM, Lang PJ. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9] [Medline: 7962581]
31. OFFICIAL Fortnite Battle Royale Menu Music - Rock Version. Addoam. URL: https://www.youtube.com/watch?v=2q-k7ScMs0k [accessed 2024-06-21]
32. Marshmello - Alone (Official Music Video). Marshmello. URL: https://www.youtube.com/watch?v=ALZHF5UqnU4 [accessed 2024-06-21]
33. Free Fire New EPIC Theme Song. FreeFire Original. URL: https://www.youtube.com/watch?v=oCBMY0MSiWA [accessed 2024-06-21]
34. Herobrine's Life (Instrumental). Abtmelody - Topic. URL: https://www.youtube.com/watch?v=Qk1FDPOP8ys [accessed 2024-06-21]
35. MC Kevinho - Olha a Explosão (KondZilla) | Official Music Video. Canal KondZilla. URL: https://www.youtube.com/watch?v=3yd_eoMOvqk [accessed 2024-06-21]
36. Rueda MR, Fan J, McCandliss BD, Halparin JD, Gruber DB, Lercari LP, et al. Development of attentional networks in childhood. Neuropsychologia 2004;42(8):1029-1040. [doi: 10.1016/j.neuropsychologia.2003.12.012] [Medline: 15093142]
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]
38. Posner MI, Petersen SE. The attention system of the human brain. Annu Rev Neurosci 1990;13(1):25-42. [doi: 10.1146/annurev.ne.13.030190.000325] [Medline: 2183676]
39. Fan J, McCandliss BD, Sommer T, Raz A, Posner MI. Testing the efficiency and independence of attentional networks. J Cogn Neurosci 2002 Apr 01;14(3):340-347. [doi: 10.1162/089892902317361886] [Medline: 11970796]
40. Fan J, Wu Y, Fossella JA, Posner MI. Assessing the heritability of attentional networks. BMC Neurosci 2001;2(1):14 [FREE Full text] [doi: 10.1186/1471-2202-2-14] [Medline: 11580865]
41. Conners CK. Conners' Continuous Performance Test for Windows. Toronto, ON: MultiHealth; 2002.
42. Conners C. Conners' Continuous Performance Test II. CPT II. North Tonawanda, NY: Multi-Health Systems; 2004.
43. Dados com Base no Levantamento Sócio Econômico IBOPE. Associação Brasileira de Empresas de Pesquisa (ABEP). 2015. URL: http://www.abep.org/new/criterioBrasil.aspx [accessed 2024-03-23]
44. Markham WA, Spencer N. Factors that mediate the relationships between household socio-economic status and childhood Attention Deficit Hyperactivity Disorder (ADHD) in children and adolescents: A systematic review. PLoS One 2022;17(3):e0262988 [FREE Full text] [doi: 10.1371/journal.pone.0262988] [Medline: 35231056]
45. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004 Mar;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
46. Johnson KA, Robertson IH, Barry E, Mulligan A, Dáibhis A, Daly M, et al. Impaired conflict resolution and alerting in children with ADHD: evidence from the Attention Network Task (ANT). J Child Psychol Psychiatry 2008 Dec;49(12):1339-1347. [doi: 10.1111/j.1469-7610.2008.01936.x] [Medline: 19120713]
47. de Souza Almeida R, Faria-Jr A, Klein RM. On the origins and evolution of the Attention Network Tests. Neurosci Biobehav Rev 2021 Jul;126:560-572. [doi: 10.1016/j.neubiorev.2021.02.028] [Medline: 33766674]
48. Hallam S, Price J, Katsarou G. The effects of background music on primary school pupils' task performance. Educational Studies 2010 Jul;28(2):111-122. [doi: 10.1080/03055690220124551]
49. Ilieva IP, Farah MJ. Attention, motivation, and study habits in users of unprescribed ADHD medication. J Atten Disord 2019 Jan;23(2):149-162. [doi: 10.1177/1087054715591849] [Medline: 26290484]
50. Smith ZR, Langberg JM. Review of the evidence for motivation deficits in youth with ADHD and their association with functional outcomes. Clin Child Fam Psychol Rev 2018 Dec;21(4):500-526. [doi: 10.1007/s10567-018-0268-3] [Medline: 30141121]
51. Morsink S, Sonuga-Barke E, Van der Oord S, Van Dessel J, Lemiere J, Danckaerts M. Task-related motivation and academic achievement in children and adolescents with ADHD. Eur Child Adolesc Psychiatry 2021 Jan;30(1):131-141. [doi: 10.1007/s00787-020-01494-8] [Medline: 32157390]
52. Skalski S, Pochwatko G, Balas R. Impact of motivation on selected aspects of attention in children with ADHD. Child Psychiatry Hum Dev 2021 Aug;52(4):586-595 [FREE Full text] [doi: 10.1007/s10578-020-01042-0] [Medline: 32816140]
53. Smith ZR, Langberg JM, Cusick CN, Green CD, Becker SP. Academic motivation deficits in adolescents with ADHD and associations with academic functioning. J Abnorm Child Psychol 2020 Feb;48(2):237-249. [doi: 10.1007/s10802-019-00601-x] [Medline: 31741133]
54. Ventouri E. ADHD and learning motivations. OALib 2020;7(8):1-28. [doi: 10.4236/oalib.1106594]
55. Shih YN, Huang HR, Chiang HY. Background music: effects on attention performance. Work 2012;42(4):573-578. [doi: 10.3233/wor-2012-1410]

56. Shih YN, Chien WH, Chiang HS. Elucidating the relationship between work attention performance and emotions arising from listening to music. Work 2016 Oct 17;55(2):489-494. [doi: 10.3233/wor-162408]
57. Franco F, Swaine JS, Israni S, Zaborowska KA, Kaloko F, Kesavarajan I, et al. Affect-matching music improves cognitive performance in adults and young children for both positive and negative emotions. Psychol Music 2014 Nov 13;42(6):869-887. [doi: 10.1177/0305735614548500]
58. Vázquez-Marrufo M, García-Valdecasas Colell M, Galvao-Carmona A, Sarrias-Arrabal E, Tirapu-Ustárroz J. [The Attention Network Test in the study of cognitive impairment of ADHD patients]. Rev Neurol 2019 Nov 16;69(10):423-432 [FREE Full text] [doi: 10.33588/rn.6910.2019202] [Medline: 31713229]

## Abbreviations

**ADHD:** attention deficit/hyperactivity disorder
**ANT:** Attention Network Test
**CBCL/6-18:** Child Behavior Checklist for Ages 6-18
**CCEB:** 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

*Edited by T de Azevedo Cardoso; submitted 23.10.23; peer-reviewed by C Kerson, V Prasad; comments to author 09.03.24; revised version received 30.03.24; accepted 29.04.24; published 18.07.24.*

*Please cite as:*
*Mendes CG, de Paula JJ, Miranda DM*
*Effects of Background Music on Attentional Networks of Children With and Without Attention Deficit/Hyperactivity Disorder: Case Control Experimental Study*
*Interact J Med Res 2024;13:e53869*
*URL: https://www.i-jmr.org/2024/1/e53869*
*doi:10.2196/53869*
*PMID:*

Case Report

# Treating Spontaneous Pneumothorax Using an Innovative Surgical Technique Called Capnodissection Pleurectomy: Case Report

Ghaith Qsous[1], MRCS, MD; Prashanth Ramaraj[1,2], MSc, MRCS; Sanjeet Singh Avtaar Singh[1], MRCS, PhD; Philip Herd[3], MBChB; Nayandra Runveer Sooraj[3], MMed, FCA (SA); Malcolm Brodie Will[1], FRCS (CTh), PhD

[1]Department of Cardiothoracic Surgery, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom

[2]Imperial College School of Medicine, Imperial College London, London, United Kingdom

[3]Department 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.

***(Interact J Med Res 2024;13:e54497)*** doi:10.2196/54497

**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 $CO_2$ insufflation for capnothorax during VATS or robot-assisted thoracoscopic 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 - ijmr_v13i1e54497_app1.mp4 ]

### References

1. Louw EH, Shaw JA, Koegelenberg CFN. New insights into spontaneous pneumothorax: a review. Afr J Thorac Crit Care Med 2021 Mar 09;27(1):10.7196/AJTCCM.2021.v27i1.054. [doi: 10.7196/ajtccm.2021.v27i1.054] [Medline: 34240041]
2. Ghisalberti M, Guerrera F, de Vico A, Bertolaccini L, de Palma A, Fiorelli A, et al. Age and clinical presentation for primary spontaneous pneumothorax. Heart Lung Circ 2020 Nov;29(11):1648-1655. [doi: 10.1016/j.hlc.2020.05.107] [Medline: 32718900]
3. Nava GW, Walker SP. Management of the secondary spontaneous pneumothorax: current guidance, controversies, and recent advances. J Clin Med 2022 Feb 22;11(5):1173 [FREE Full text] [doi: 10.3390/jcm11051173] [Medline: 35268264]
4. Roberts ME, Rahman NM, Maskell NA, Bibby AC, Blyth KG, Corcoran JP, et al. British Thoracic Society guideline for pleural disease. Thorax 2023 Nov 08;78(11):1143-1156. [doi: 10.1136/thorax-2023-220304] [Medline: 37553157]
5. Ocakcioglu I, Kupeli M. Surgical treatment of spontaneous pneumothorax: pleural abrasion or pleurectomy? Surg Laparosc Endosc Percutan Tech 2019 Feb;29(1):58-63. [doi: 10.1097/SLE.0000000000000595] [Medline: 30499890]
6. Gallego-Poveda J, Guerra NC, Carvalheiro C, Ferreira H, Sena A, Junqueira N, et al. Use of CO in video assisted thoracic surgery and single-lumen endotracheal tube-a new less invasive approach. J Thorac Dis 2017 Apr;9(4):903-906 [FREE Full text] [doi: 10.21037/jtd.2017.01.53] [Medline: 28523135]
7. Brock H, Rieger R, Gabriel C, Pölz W, Moosbauer W, Necek S. Haemodynamic changes during thoracoscopic surgery the effects of one-lung ventilation compared with carbon dioxide insufflation. Anaesthesia 2000 Jan;55(1):10-16 [FREE Full text] [doi: 10.1046/j.1365-2044.2000.01123.x] [Medline: 10594427]

8. Körner H, Andersen KS, Stangeland L, Ellingsen I, Engedal H. Surgical treatment of spontaneous pneumothorax by wedge resection without pleurodesis or pleurectomy. Eur J Cardiothorac Surg 1996;10(8):656-659. [doi: 10.1016/s1010-7940(96)80381-2] [Medline: 8875174]
9. Chang J, Ratnaraj V, Fu V, Jiang M, Peri V, Nguyenhuy M, et al. Pleural abrasion versus apical pleurectomy for primary spontaneous pneumothorax: a systematic review and meta-analysis. J Cardiothorac Surg 2023 Apr 06;18(1):105 [FREE Full text] [doi: 10.1186/s13019-023-02207-3] [Medline: 37024894]
10. Hallifax RJ, Yousuf A, Jones HE, Corcoran JP, Psallidas I, Rahman NM. Effectiveness of chemical pleurodesis in spontaneous pneumothorax recurrence prevention: a systematic review. Thorax 2017 Dec 01;72(12):1121-1131 [FREE Full text] [doi: 10.1136/thoraxjnl-2015-207967] [Medline: 27803156]
11. Dai J, Liu M, Liu X, Li J, Jin K, Chen L, et al. Carbon dioxide blower facilitates visceral pleurectomy in malignant pleural mesothelioma. Ann Thorac Surg 2022 Jul;114(1):e71-e74. [doi: 10.1016/j.athoracsur.2021.09.046] [Medline: 34715083]
12. Tran DTT, Badner NH, Nicolaou G, Sischek W. Arterial pCO2 changes during thoracoscopic surgery with CO2 insufflation and one lung ventilation. HSR Proc Intensive Care Cardiovasc Anesth 2010;2(3):191-197 [FREE Full text] [Medline: 23439224]
13. van den Heuvel MM, Smit HJ, Barbierato SB, Havenith CE, Beelen RH, Postmus PE. Talc-induced inflammation in the pleural cavity. Eur Respir J 1998 Dec 01;12(6):1419-1423 [FREE Full text] [doi: 10.1183/09031936.98.12061419] [Medline: 9877502]

## Abbreviations

**BTS:** British Thoracic Society
**$CO_2$:** carbon dioxide
**SP:** spontaneous pneumothorax
**VATS:** video-assisted thorascopic surgery

*Edited by T de Azevedo Cardoso; submitted 12.11.23; peer-reviewed by UK Chalwadi, A Chambers; comments to author 16.02.24; revised version received 02.04.24; accepted 27.05.24; published 21.06.24.*

*Please cite as:*
*Qsous G, Ramaraj P, Avtaar Singh SS, Herd P, Sooraj NR, Will MB*
*Treating Spontaneous Pneumothorax Using an Innovative Surgical Technique Called Capnodissection Pleurectomy: Case Report*
*Interact J Med Res 2024;13:e54497*
*URL: https://www.i-jmr.org/2024/1/e54497*
*doi:10.2196/54497*
*PMID:*

Case Report

# Radiological Progression of Degenerative Cervical Myelopathy in a Clinically Stable Patient: Case Report

Rishi Umeria[1], MA, MB BChir; Oliver Mowforth[1], MA, MB BChir, MSt, MRCS; Munashe Veremu[1], BSc, MBBS; Benjamin Davies[1], MBChB, BSc, MRCS; Mark Kotter[1], 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.

***(Interact J Med Res 2024;13:e48212)*** doi:10.2196/48212

**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.

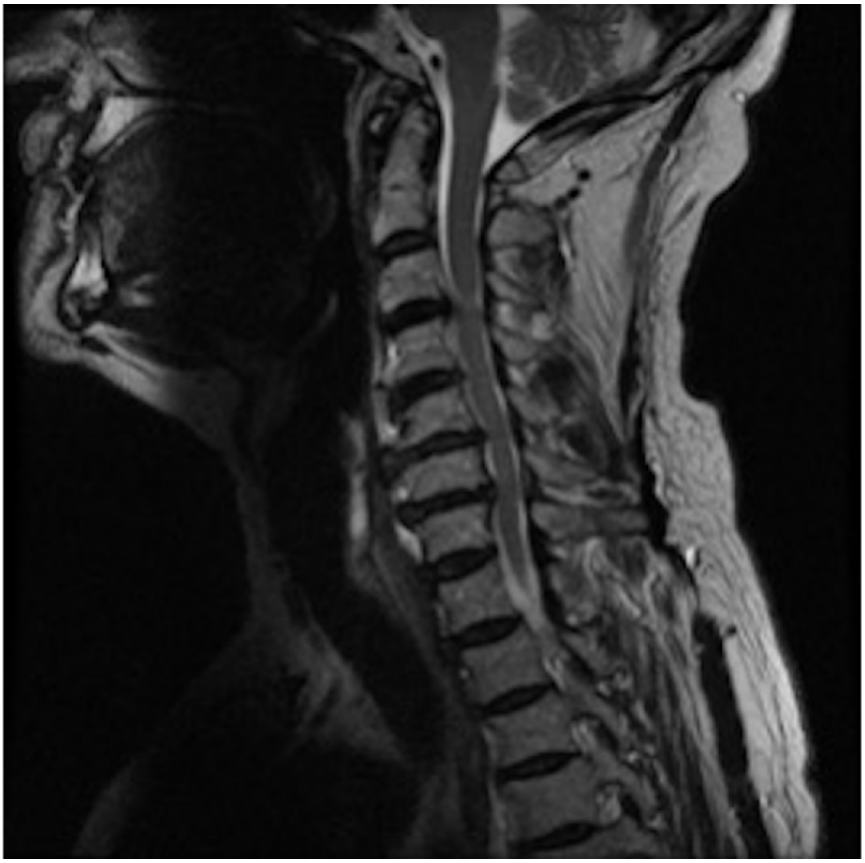


**Figure 2.** T2-weighted sagittal MRI demonstrating the progression of multilevel degenerative changes in the cervical spine, including disc degeneration at C3-C7.

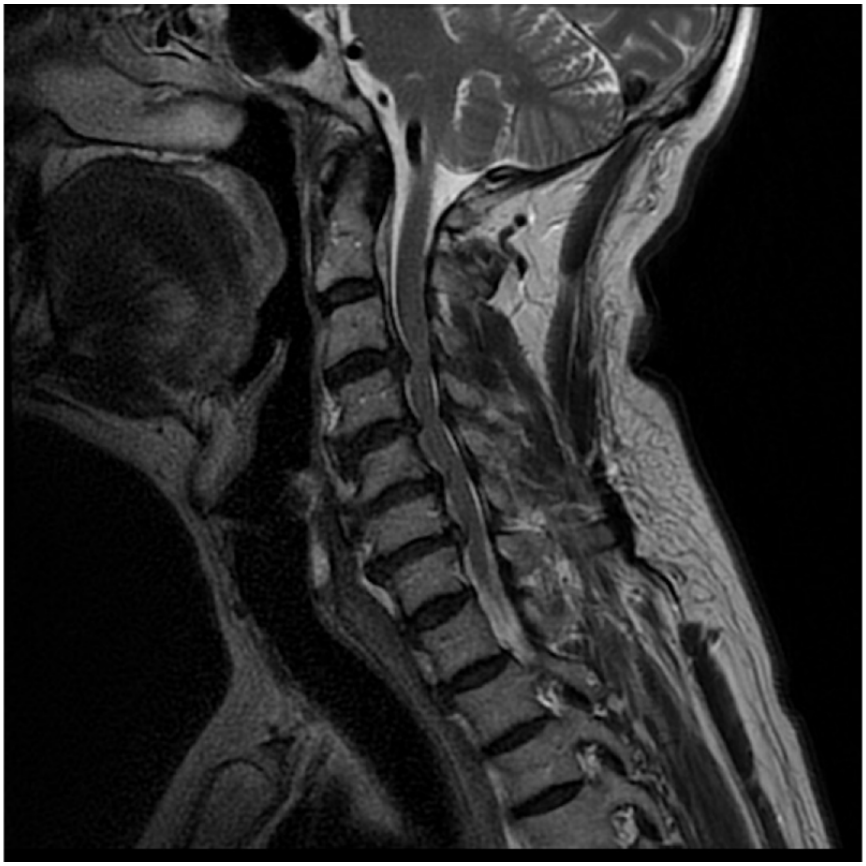
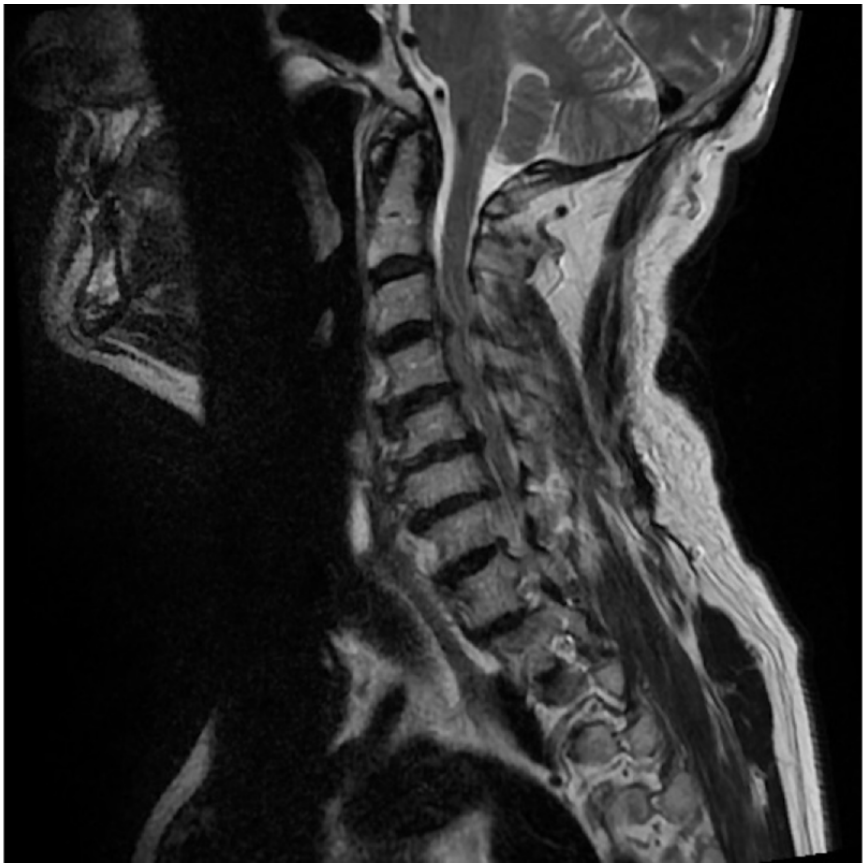


**Figure 3.** T2-weighted sagittal magnetic resonance imaging showing significant radiological worsening of cervical cord compression at C3-C7 levels despite unchanged symptoms and modified Japanese Orthopaedic Association clinical severity score.



### 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 $B_{12}$ 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

1. Davies BM, Mowforth OD, Smith EK, Kotter MR. Degenerative cervical myelopathy. BMJ 2018;360:k186 [FREE Full text] [doi: 10.1136/bmj.k186] [Medline: 29472200]
2. Smith SS, Stewart ME, Davies BM, Kotter MRN. The prevalence of asymptomatic and symptomatic spinal cord compression on magnetic resonance imaging: a systematic review and meta-analysis. Global Spine J 2021;11(4):597-607 [FREE Full text] [doi: 10.1177/2192568220934496] [Medline: 32677521]
3. Fehlings MG, Wilson JR, Yoon ST, Rhee JM, Shamji MF, Lawrence BD. Symptomatic progression of cervical myelopathy and the role of nonsurgical management: a consensus statement. Spine (Phila Pa 1976) 2013;38(22 Suppl 1):S19-S20 [FREE Full text] [doi: 10.1097/BRS.0b013e3182a7f4de] [Medline: 23963011]
4. Witiw CD, Mathieu F, Nouri A, Fehlings MG. Clinico-radiographic discordance: an evidence-based commentary on the management of degenerative cervical spinal cord compression in the absence of symptoms or with only mild symptoms of myelopathy. Global Spine J 2018;8(5):527-534 [FREE Full text] [doi: 10.1177/2192568217745519] [Medline: 30258760]

5. Harrop JS, Naroji S, Maltenfort M, Anderson DG, Albert T, Ratliff JK, et al. Cervical myelopathy: a clinical and radiographic evaluation and correlation to cervical spondylotic myelopathy. Spine (Phila Pa 1976) 2010;35(6):620-624. [doi: 10.1097/BRS.0b013e3181b723af] [Medline: 20150835]
6. Nemani VM, Kim HJ, Piyaskulkaew C, Nguyen JT, Riew KD. Correlation of cord signal change with physical examination findings in patients with cervical myelopathy. Spine (Phila Pa 1976) 2015;40(1):6-10. [doi: 10.1097/BRS.0000000000000659] [Medline: 25341986]
7. Milligan J, Ryan K, Fehlings M, Bauman C. Degenerative cervical myelopathy: diagnosis and management in primary care. Can Fam Physician 2019;65(9):619-624 [FREE Full text] [Medline: 31515310]
8. Fehlings MG, Tetreault LA, Riew KD, Middleton JW, Aarabi B, Arnold PM, et al. A clinical practice guideline for the management of patients with degenerative cervical myelopathy: recommendations for patients with mild, moderate, and severe disease and nonmyelopathic patients with evidence of cord compression. Global Spine J 2017;7(Suppl 3):70S-83S [FREE Full text] [doi: 10.1177/2192568217701914] [Medline: 29164035]
9. Akter F, Yu X, Qin X, Yao S, Nikrouz P, Syed YA, et al. The pathophysiology of degenerative cervical myelopathy and the physiology of recovery following decompression. Front Neurosci 2020;14:138 [FREE Full text] [doi: 10.3389/fnins.2020.00138] [Medline: 32425740]
10. Kopjar B, Tetreault L, Kalsi-Ryan S, Fehlings M. Psychometric properties of the modified Japanese Orthopaedic Association scale in patients with cervical spondylotic myelopathy. Spine (Phila Pa 1976) 2015;40(1):E23-E28 [FREE Full text] [doi: 10.1097/BRS.0000000000000648] [Medline: 25341993]
11. Sangeorzan I, Antonacci G, Martin A, Grodzinski B, Zipser CM, Murphy RKJ, et al. Toward shared decision-making in degenerative cervical myelopathy: protocol for a mixed methods study. JMIR Res Protoc 2023;12:e46809 [FREE Full text] [doi: 10.2196/46809] [Medline: 37812472]
12. Davies BM, Mowforth O, Gharooni AA, Tetreault L, Nouri A, Dhillon RS, et al. A new framework for investigating the biological basis of degenerative cervical myelopathy [AO spine RECODE-DCM research priority number 5]: mechanical stress, vulnerability and time. Global Spine J 2022;12(Suppl 1):78S-96S [FREE Full text] [doi: 10.1177/21925682211057546] [Medline: 35174728]
13. Wang ZC, Hou XW, Shao J, Ji YJ, Li L, Zhou Q, et al. HIF-1α polymorphism in the susceptibility of cervical spondylotic myelopathy and its outcome after anterior cervical corpectomy and fusion treatment. PLoS One 2014;9(11):e110862 [FREE Full text] [doi: 10.1371/journal.pone.0110862] [Medline: 25401740]
14. Badhiwala JH, Ahuja CS, Akbar MA, Witiw CD, Nassiri F, Furlan JC, et al. Degenerative cervical myelopathy—update and future directions. Nat Rev Neurol 2020;16(2):108-124. [doi: 10.1038/s41582-019-0303-0] [Medline: 31974455]
15. Senter HJ, Venes JL. Loss of autoregulation and posttraumatic ischemia following experimental spinal cord trauma. J Neurosurg 1979;50(2):198-206. [doi: 10.3171/jns.1979.50.2.0198] [Medline: 430132]
16. Saadoun S, Papadopoulos MC. Targeted perfusion therapy in spinal cord trauma. Neurotherapeutics 2020;17(2):511-521 [FREE Full text] [doi: 10.1007/s13311-019-00820-6] [Medline: 31916236]
17. Srihari G, Shukla D, Devi BI, Sathyaprabha TN. Subclinical autonomic nervous system dysfunction in compressive cervical myelopathy. Spine (Phila Pa 1976) 2011;36(8):654-659. [doi: 10.1097/BRS.0b013e3181dc9eb2] [Medline: 21178837]
18. Montine TJ, Cholerton BA, Corrada MM, Edland SD, Flanagan ME, Hemmy LS, et al. Concepts for brain aging: resistance, resilience, reserve, and compensation. Alzheimers Res Ther 2019;11(1):22 [FREE Full text] [doi: 10.1186/s13195-019-0479-y] [Medline: 30857563]
19. Zdunczyk A, Schwarzer V, Mikhailov M, Bagley B, Rosenstock T, Picht T, et al. The corticospinal reserve capacity: reorganization of motor area and excitability as a novel pathophysiological concept in cervical myelopathy. Neurosurgery 2018;83(4):810-818. [doi: 10.1093/neuros/nyx437] [Medline: 29165642]


## Abbreviations

**DCM:** degenerative cervical myelopathy
**mJOA:** modified Japanese Orthopaedic Association
**MRI:** magnetic resonance imaging


*Edited by T de Azevedo Cardoso; submitted 15.04.23; peer-reviewed by Y Wei, H Alsalhi; comments to author 12.12.23; revised version received 23.01.24; accepted 02.04.24; published 27.06.24.*

*Please cite as:*
*Umeria R, Mowforth O, Veremu M, Davies B, Kotter M*
*Radiological Progression of Degenerative Cervical Myelopathy in a Clinically Stable Patient: Case Report*
*Interact J Med Res 2024;13:e48212*
*URL: https://www.i-jmr.org/2024/1/e48212*
*doi:10.2196/48212*
*PMID:*

Research Letter

# Visual "Scrollytelling": Mapping Aquatic Selfie-Related Incidents in Australia

Samuel Cornell[1], BSc, MSc; Amy E Peden[1], BA, MPP, PhD

School of Population Health, University of New South Wales, Sydney, Australia

**Corresponding Author:**
Samuel Cornell, BSc, MSc
School of Population Health
University of New South Wales
High St
Kensington
Sydney, 2052
Australia
Phone: 61 (2) 9065 38
Email: s.cornell@unsw.edu.au

***(Interact J Med Res 2024;13:e53067)*** doi:10.2196/53067

**KEYWORDS**

selfie; map; social media; selfies; scrollama; JavaScript; scrollytelling; Mapbox; incidence; incidents; incident; fatality; fatalities; injury; injuries; retrieval; prevalence; image; images; photo; photos; photograph; photographs; prevalence; Australia; emergency; visualization; visualizations; interactive; location; geography; geographic; geographical; spatial; artificial intelligence; longitude; latitude; visual representation; visual representations

## 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 scrollable visual 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.

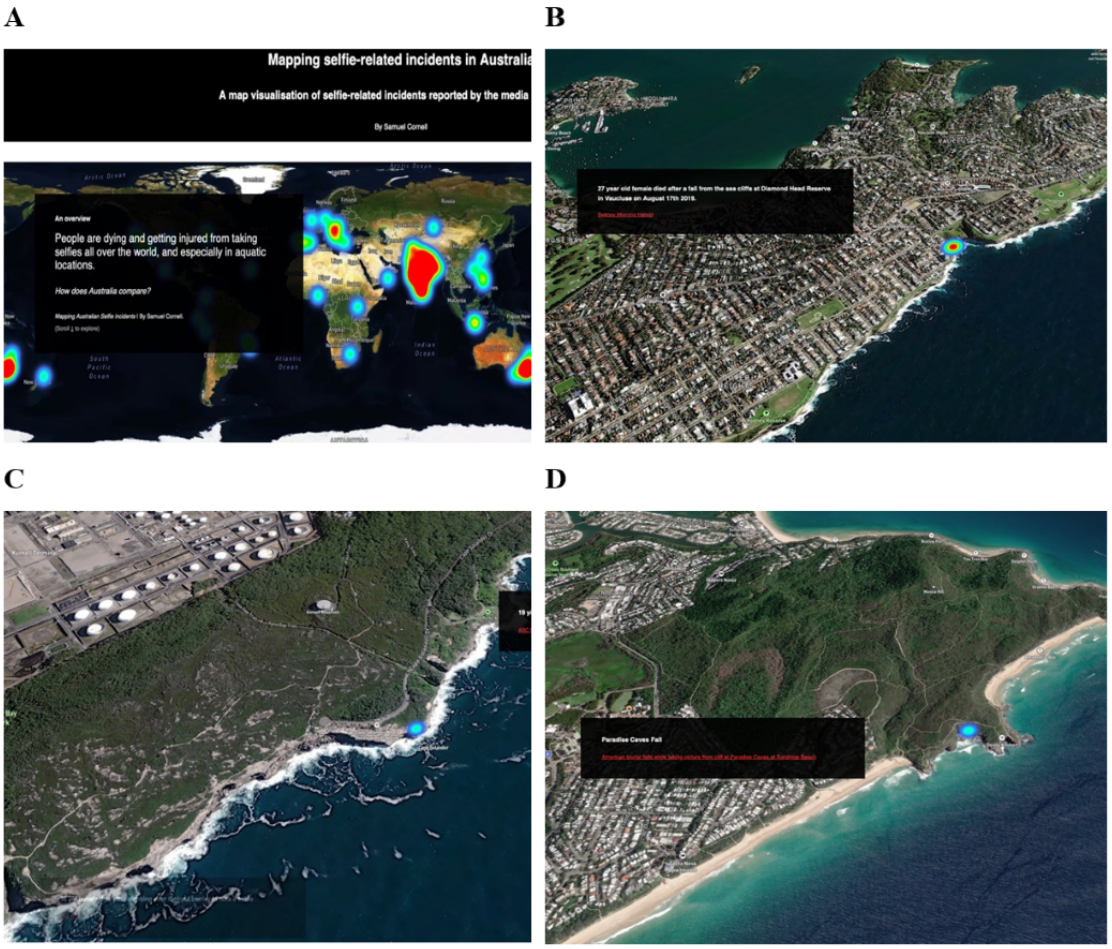
### Ethical Considerations

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.

**Figure 1.** (A) Image taken from the web-based site heat map of worldwide selfie-related incidents. The image provides a worldwide overview of incidents based on the obtained media data used in this study. (B), (C), and (D) Example images acquired from the web-based site. These images illustrate the scrolling story of the heat map focusing on a location that has seen selfie-related incidents. Each incident is indicated in a "chapter," which provides a description of the incident in that location and a link to the corresponding news report. Images were acquired from Mapbox [7] and OpenStreetMap [8]. OpenStreetMap is licensed under the Open Data Commons Open Database License [9].




## 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.
[DOCX File , 17 KB - ijmr_v13i1e53067_app1.docx ]

## References

1. Cornell S, Brander RW, Peden AE. Preventing selfie-related incidents: taking a public health approach to reduce unnecessary burden on emergency medicine services. Emerg Med Australas 2023 Aug 19;35(4):691-693. [doi: 10.1111/1742-6723.14219] [Medline: 37454361]
2. List of selfie-related injuries and deaths. Wikipedia. URL: https://tinyurl.com/4xfwsfce [accessed 2023-01-11]
3. Location Helper. URL: https://labs.mapbox.com/location-helper/#3/4.48/49.2 [accessed 2023-09-06]
4. Design beautiful custom maps with Mapbox Studio. Mapbox. URL: https://www.mapbox.com/mapbox-studio [accessed 2023-09-06]
5. mapbox / storytelling. GitHub. URL: https://github.com/mapbox/storytelling [accessed 2023-09-06]
6. Cornell S. Mapping selfie and social media-related incidents in Australia. URL: https://samuelcornell.github.io/selfiemap/source/index.html [accessed 2023-09-25]
7. Mapbox. URL: https://www.mapbox.com/ [accessed 2024-05-22]
8. OpenStreetMap. URL: https://www.openstreetmap.org/ [accessed 2024-05-22]
9. Open Data Commons Open Database License. URL: https://www.openstreetmap.org/copyright [accessed 2024-05-22]
10. samuelcornell / selfiemap. GitHub. URL: https://github.com/samuelcornell/selfiemap/ [accessed 2024-05-09]

*Edited by T de Azevedo Cardoso; submitted 25.09.23; peer-reviewed by N Auger, S Kang-Auger, J Luo; comments to author 06.12.23; revised version received 14.12.23; accepted 27.02.24; published 23.05.24.*

*Please cite as:*
*Cornell S, Peden AE*
*Visual "Scrollytelling": Mapping Aquatic Selfie-Related Incidents in Australia*
*Interact J Med Res 2024;13:e53067*
*URL: https://www.i-jmr.org/2024/1/e53067*
*doi:10.2196/53067*
*PMID:38781002*

Commentary

# Supporting the Mind in Space: Psychological Tools for Long-Duration Missions

Francesco Pagnini[1], PsyD, PhD

Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

**Corresponding Author:**
Francesco Pagnini, PsyD, PhD
Department of Psychology
Università Cattolica del Sacro Cuore
L.go Gemelli 1
Milan, 20123
Italy
Email: Francesco.Pagnini@unicatt.it

**Related Article:**

Comment on: https://www.i-jmr.org/2024/1/e58803

## Abstract

The psychological well-being of astronauts is becoming just as vital as their physical and technical readiness as space missions extend into deep space. Long-duration missions pose unique challenges, such as isolation, confinement, communication delays, and microgravity, which can significantly affect mental health and cognitive performance. This commentary discusses the need for innovative mental health support systems, including automated psychotherapy, as well as Earth-based training methods like mindfulness and relaxation techniques, to address the psychological demands of space travel. By integrating these approaches into pre-mission preparation and in-flight routines, astronauts can develop self-regulation strategies to manage stress, improve focus, and enhance emotional resilience. Automated psychotherapy available 24-7 provides real-time confidential support when communication with Earth is delayed. As space exploration moves forward, the success of missions will depend not only on technological advancements but also on the development of psychological countermeasures that prioritize mental health alongside physical well-being. This paper emphasizes the importance of continued research and collaboration to refine and test these tools in analog environments, ensuring astronauts are mentally and emotionally prepared for the challenges of space.

***(Interact J Med Res 2024;13:e66626)*** doi:10.2196/66626

**KEYWORDS**

space psychology; astronauts; psychotherapy; isolated and confined environment; mindfulness; relaxation; mind-body

## Introduction

The manuscript “Automated Psychotherapy in a Spaceflight Environment: Advantages, Drawbacks, and Unknowns,” written by Smith [1], explores the potential of automated psychotherapy to address the mental health challenges faced by astronauts during space missions. This technique could be highly relevant for human space exploration, particularly in overcoming delayed communication issues in long-distance missions. However, further research is needed in this field, which represents a new frontier in both space exploration and mental health care.

## Psychological Aspects of Space Exploration

As humanity pushes further into space, the psychological well-being of astronauts becomes as critical as the technical aspects of a mission [1]. Long-duration missions, including those on future Mars and lunar habitats, present unique challenges beyond just the physical demands [2]. Astronauts face months, even years, of isolation, confinement, communication delays, and microgravity in an extreme environment that is, from multiple perspectives, “hostile” to human life. These factors can impact mental health; cognitive performance; and, ultimately, mission success [3].

Astronauts are highly trained individuals, but they are still human. Prolonged isolation and distance from Earth, coupled with the stress of living and working in space, can lead to undesirable psychological reactions and reduced cognitive function [4,5]. As missions become longer and more distant, space agencies need to consider not only the physical but also the psychological needs of astronauts to ensure that they can function at their best.

## Psychological Support in Space

Providing psychological support during space missions is not a new concept [6], but as missions extend into deep space, the challenges become more complex. Traditionally, astronauts have had access to Earth-based psychologists via private communications. Private consultations and ongoing support can help manage interpersonal tensions, alleviate feelings of isolation, and address stress in real time. While this has worked well for missions in low earth orbit, such as those aboard the International Space Station, long communication delays in deep space are expected to make this system less effective [2]. For example, a Mars mission could experience a communication delay of up to 30 minutes each way. In such cases, astronauts need tools that allow for real-time mental health support, including the possibility of an automated psychotherapy system. This is also where pre-mission psychological training and on-demand support systems come into play. These include resilience training, team dynamics workshops, and tailored psychological tools that astronauts can access during the mission.

## Mindfulness and Relaxation: Key Tools for Managing Stress

One promising approach to managing stress during long missions is through mindfulness and relaxation techniques [7]. These practices have been shown to substantially reduce stress, enhance focus, and improve emotional regulation, making them valuable tools for astronauts facing long-term isolation and confinement. Mindfulness encourages individuals to stay present and manage stress without judgment, which can be particularly helpful in an environment where stressors are constant and inevitable. Astronauts trained in mindfulness techniques are expected, according to subject matter experts, to handle stress more effectively, maintaining mental clarity and decision-making abilities even in high-pressure situations [8]. Recently, mindfulness disposition has been demonstrated as a protective factor against stress in an analog environment, the Concordia base in Antarctica [9].

Relaxation training, such as diaphragmatic breathing and progressive muscle relaxation, can also play an important role. These techniques can reduce physiological stress, promote better sleep, and help astronauts manage their emotional responses to challenging situations. Given the stressful and sometimes monotonous environment of space, these practices offer simple yet powerful ways to maintain both mental and physical well-being [7].

By integrating mindfulness and relaxation into astronauts' training, space agencies can equip them with effective self-regulation tools to help manage the ongoing psychological demands of space missions. These methods are not only noninvasive but also easy to practice, making them ideal for space, where time and resources are limited.

## The Role of Automated Psychotherapy in Space

With communication delays limiting real-time support from Earth, automated psychotherapy offers a crucial solution for mental health care during space missions. These systems can provide astronauts with immediate access to cognitive behavioral therapy and other interventions tailored to their specific needs and the unique stresses of space travel. One of the most compelling aspects of Smith's [1] manuscript is the focus on how automated psychotherapy could overcome the challenges of delayed communication in deep space missions. This is particularly relevant as human space exploration moves toward Mars and beyond, where real-time communication with Earth will no longer be possible. Smith's [1] work provides a strong foundation for understanding the psychological complexities of long-duration missions, particularly the potential for automated systems to fill gaps in mental health support when traditional Earth-based therapy is not feasible.

Automated psychotherapy tools also offer additional advantages. They are available 24-7, offering astronauts confidential on-demand support. This is particularly useful for addressing sensitive issues that astronauts might not feel comfortable discussing with crewmates or mission control. These systems can guide astronauts through therapeutic exercises, helping them manage anxiety, stress, or feelings of isolation as they arise. Furthermore, automated systems can be integrated with other mental health tools, such as mindfulness and relaxation practices, creating a comprehensive mental health support network. This combination of tools ensures that astronauts can address both immediate psychological needs and long-term mental health maintenance without relying on Earth-based professionals. As these systems are refined, they have the potential to become a critical component of mental health care for long-duration missions.

## Conclusion

As space exploration advances into longer, more isolated missions, the psychological and human aspects of space travel become increasingly critical. The challenges of isolation, stress, and confinement in space demand innovative solutions that go beyond traditional mental health support [2].

The integration of automated psychotherapy, as well as newly developed training techniques (eg, mindfulness and relaxation training), offer a proactive and effective approach to safeguarding astronaut mental health. By providing astronauts with these essential tools, space agencies can ensure that crews are not only physically prepared but also mentally resilient, capable of adapting to the extreme demands of long-duration missions. The success of future space exploration, however,

will depend on our collective commitment to supporting the psychological well-being of astronauts. The development, testing, and refinement of these psychological tools in analog environments are imperative. As we advance into the next phase of space exploration, space agencies, researchers, and innovators need to work together to ensure that mental health receives the same attention as physical and technical preparation, helping astronauts thrive on their journeys into the unknown.

## Acknowledgments

This work was supported by the Italian Space Agency, through the projects RelaxPro (DC-VUM-2020-007, Mind/body practices for deep space exploration) and MINDFUL-ICE II (DC-DSR-UVS-2022-212, "A Quick Mindfulness Training for an Isolated and Confined Environment"), the European Space Agency, the French Polar Institute Paul-Émile Victor, and the Programma Nazionale di Ricerche in Antartide.

Artificial intelligence (ChatGPT) was used to spell-check and polish the text.

## Conflicts of Interest

None declared.

## References

1. Smith L. Automated Psychotherapy in a Spaceflight Environment: Advantages, Drawbacks, and Unknowns. Interactive J Med Res 2024;13:e58803. [doi: 10.2196/58803]
2. Pagnini F, Manzey D, Rosnet E, Ferravante D, White O, Smith N. Human behavior and performance in deep space exploration: next challenges and research gaps. NPJ Microgravity 2023 Mar 30;9(1):27. [doi: 10.1038/s41526-023-00270-7] [Medline: 36997549]
3. Nicolas M, Martinent G, Palinkas L, Suedfeld P. Dynamics of stress and recovery and relationships with perceived environmental mastery in extreme environments. J Environ Psychol 2022 Oct;83:101853. [doi: 10.1016/j.jenvp.2022.101853]
4. Stahn AC, Bucher D, Zu Eulenburg P, Denise P, Smith N, Pagnini F, et al. Paving the way to better understand the effects of prolonged spaceflight on operational performance and its neural bases. NPJ Microgravity 2023 Jul 31;9(1):59. [doi: 10.1038/s41526-023-00295-y] [Medline: 37524737]
5. Palinkas LA, Suedfeld P. Psychosocial issues in isolated and confined extreme environments. Neurosci Biobehav Rev 2021 Jul;126:413-429. [doi: 10.1016/j.neubiorev.2021.03.032] [Medline: 33836214]
6. Manzey D. Human missions to Mars: new psychological challenges and research issues. Acta Astronaut 2004;55(3-9):781-790. [doi: 10.1016/j.actaastro.2004.05.013] [Medline: 15806750]
7. Pagnini F, Phillips D, Bercovitz K, Langer E. Mindfulness and relaxation training for long duration spaceflight: evidences from analog environments and military settings. Acta Astronautica 2019 Dec;165:1-8. [doi: 10.1016/j.actaastro.2019.07.036]
8. Pagnini F, Phillips D, Langer E. Mindfulness, recreation, and relaxation for exploration missions. 2017 Presented at: NASA Human Research Program Investigators' Workshop. A New Dawn: Enabling Human Space Exploration; January 23-26, 2017; Galveston, TX.
9. Pagnini F, Thoolen S, Smith N, Van Ombergen A, Grosso F, Langer E, et al. Mindfulness disposition as a protective factor against stress in Antarctica: a potential countermeasure for long-duration spaceflight? J Environ Psychol 2024 Mar;94:102254. [doi: 10.1016/j.jenvp.2024.102254]

*Edited by T de Azevedo Cardoso; submitted 18.09.24; this is a non–peer-reviewed article;accepted 20.09.24; published 09.10.24.*

*Please cite as:*
*Pagnini F*
*Supporting the Mind in Space: Psychological Tools for Long-Duration Missions*
*Interact J Med Res 2024;13:e66626*
*URL: https://www.i-jmr.org/2024/1/e66626*
*doi:10.2196/66626*
*PMID:39382951*

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org